SELECTION PROCEDURES IN POPULATIONS

by

SUSANNE UNDERWOOD

B.Sc. (London)

DIP.ANIM.GENET. (Edinburgh)

Thesis presented for the Degree of Doctor of Philosophy of the University of Edinburgh in the Faculty of Science

December 1971

CONTENTS

SUMMARY

A theoretical study has been made of selection operating in finite populations of different structures. The aim has been to investigate ways of using selection procedures involving sub-populations in order to maximize the limit to selection. Most of the results presented have been obtained by Monte Carlo simulation, although algebraic analysis has been used where possible.

Various genetic models have been considered:

(i) Single or independent locus models. Existing theory has been applied and extended for loci of both additive and non-additive effect.

(ii) Multi-locus models. Much of the work has been under a simple two locus additive model with some extensions to the case of more than two loci.

Two basic situations have been considered with respect to population structure taking:

(i) a single base population in linkage equilibrium and

(ii) two distinct base populations each of which were in linkage equilibrium.

For both cases the response at the limit for simple mass selection in a population of size N (taken from the original base population in case (i) and from a cross of the two populations in case (ii)) has been compared with the ultimate response obtained using a system of population sub-division and recrossing in a population always of total size N.

Since such systems are likely to involve linkage disequilibrium, under the multi-locus models, the importance of this has initially been studied. It has been found that the negative linkage disequilibrium which may be produced by crossing distinct populations can have a considerable effect on response to selection. The reduction caused is greatest for those loci which have favourable alleles at low frequency or are of small selective advantage. In order to study linkage disequilibrium under a model of more than two loci it was necessary to first develop a system of measurement which took into account both the distribution of alleles between chromosomes and the order of loci on them. Two measures were derived and compared and both were found to be useful as predictors of ultimate response.

Linkage disequilibrium can be reduced by a period of relaxation after crossing and prior to selection but results presented here suggest that this is unlikely to be of any practical value.

Studies on selection from a single base population have shown that it is possible to increase the selection limit, for the single locus case at least, using some system of subdivision and recrossing, but the magnitude of the increase is extremely small and seems unlikely to be of any practical value. In general, however, population sub-division tends to give the same ultimate limit but with a reduced response at any given intermediate generation.

If two populations are to form the basis of a selection scheme then it has been shown that in general a cross taking equal proportions from each will give the highest limit. Unless considerable differences exist, in terms of favourable alleles present, between the two populations and important loci are closely linked then linkage disequilibrium will have very little effect on the response and selection in sub-lines prior to crossing will not increase the ultimate limit. However if such differences exist that the linkage disequilibrium generated by the cross is of importance then a higher limit may be attained by selecting the populations as separate sub-lines before crossing. Even so the overall rate of response will be reduced and the effect of this in intermediate generations may mean that such a selection procedure is still not feasible in economic terms.

CHAPTER I

Introduction

Population genetics theory is primarily concerned with changes in gene frequency within a population; the aim being to predict changes under all possible conditions. The most basic of these predictions is the Hardy-Weinberg Law which applies to an individual gene in a population under the following restrictions;

i) Infinite population size,

ii) Random mating,

iii) No selection,

iv) No mutation

v) No migration,

vi) Non-overlapping generations,

vii) No sex linkage.

Under these conditions the Hardy-Weinberg Law states that both gene and genotype frequencies will remain constant from generation to generation. The most difficult of the above restrictions to lift has proved to be that regarding population size. The remaining restrictions have been removed, to some extent at least, and theory has been developed to enable predictions to be made under a wide variety of conditions (Falconer, 1960).

In reality no population can be infinite in size but for large populations infinite population theory provides an acceptable approximation. Young (1966) carried out computer simulation work on selection in large populations, selecting between 100 and 800 individuals from 1000 each generation. He found that the responses observed agreed well with predictions made from infinite population

theory. However, both in evolution and artificial selection populations considerably smaller in size may be of importance. In view of this, much work has recently been carried out on changes occurring in populations of small size, with particular reference to the way in which selection operates. This work will be reviewed before going on to a consideration of the importance of structure within the finite population.

Selection in a Finite Population

The changes produced by selection operating in a population of finite size cannot be predicted exactly because of random changes produced by sampling. Therefore infinite population theory must be combined with probability theory to make predictions in populations of finite size. This problem is perhaps best approached by considering first the case of selection at a single locus.

Single locus theory

Consider a single locus at which there are two alleles segregating in a random population of effective size N, as used by Wright (1931). Let the alleles be denoted A and a with initial frequencies p and (1-p) respectively. The expected frequency of A in the next generation is then given by standard infinite population theory and is dependent upon the relative selective advantages of the two homozygous genotypes, say s, and a function of the gene frequency, say f(p), depending upon the type of gene action involved. The change in gene frequency, Δp, expected after one generation of selection can then be expressed as

$$E(\Delta p) = s.f.(p) \qquad \ldots(1)$$

The equations for Δp in an infinite population are given by Falconer (1960).

In a finite population sampling of a relatively small number of gametes from a large pool occurs. The expected change in the number of A and a gametes is given by $\Delta p.N$ and $(1-\Delta p)N$ respectively. However, by chance alone the actual number of each type of gamete sampled may differ from expectation. The magnitude of this difference in any given trial cannot be predicted but the variance in Δp can be formulated from the variance of the binomial to give

$$\sigma^2 \Delta p \simeq p(1-p)/2N \text{ approximately} \qquad \ldots(2)$$

This is the concept of "Random Drift" of gene frequency, essentially due to Fisher (1930) and Wright (1931).

When there is no selection operating the expected change in frequency of either allele is zero. The expected frequency of A in the next generation is then simply p. However, if sampling causes a change in gene frequency to p' then this becomes the expected frequency for the next generation and so on. This process continues until p reaches a value of one or zero, after which no further change can occur in the absence of mutation or migration. Such a state is termed 'fixation', with A being fixed if $p = 1$. This leads to the concept of "chance of fixation", which may be defined as the probability that a particular allele will eventually reach a frequency of one. In more practical terms this may be regarded in one of two ways (Robertson, 1960):-

i) as the proportion of equivalent loci which might be expected to be fixed for a particular allele type in any given line

ii) as the proportion of a set of replicate lines in which a

particular allele at an individual locus might be expected to be fixed.

The time until fixation is reached is itself a variable whose distribution is discussed below. When s = 0 the chance of fixation is easily obtained by considering a large number of replicate lines, each of finite size. Then, in spite of fluctuations of frequency within the lines, the mean frequency over all lines will remain constant at p, even after all the lines are fixed. Therefore, the chance of fixation in the absence of selection is simply the initial frequency in the population.

When selection is operating the situation is not nearly so simple. The systematic changes in frequency made by selection are dependent upon the actual frequency at any time and so random changes due to sampling become confounded with the directional changes due to selection. One approach to this problem, which has been widely used, is that of treating the system as a stochastic process of the Markov type. Broadly, this may be applied to systems where future changes can be said to depend on the present state but not on the past history which led up to the present state (Feller, 1950). Therefore, changes in gene frequency due to sampling and selection can be treated as Markov chains and these may be studied in several ways, for example:-

a) by simulation techniques, which will be discussed later,

b) by matrix methods, which will also be considered later,

c) by the use of partial differential equations. This method requires that the changes in gene frequency be treated as varying continuously with time and is strictly only applicable when changes

in gene frequency occur in a very large population over a very long time. However, in spite of this reservation, this method has been widely used even under situations where rapid changes occur in very small populations. The errors involved in such approximate methods will be discussed further below. Kimura (1964) has applied partial differential equation methods to a wide variety of problems; in particular, to examine the process of natural selection in finite populations (Kimura, 1955a, 1955b). In these investigations he applied the Kolmogorov forward equation to describe the process of change in gene frequency with time as

$$\frac{\partial \phi(x,t)}{t} = \frac{1}{2} \cdot \frac{\partial^2}{\partial x^2} [\sigma^2(x)\ \phi\ (x,t)] - \frac{\partial}{\partial x} [M(x)\ \phi\ (x,t)] \quad \ldots(3)$$

where $\phi\ (x,t)$ = the probability density that the gene frequency lies between x and x+ dx at the t^{th} generation given that the initial frequency is p at generation t = 0.

$\sigma^2(x)$ = the variance in the change in gene frequency due to random drift.

= $x\ (1-x)/2N$

$M(x)$ = the mean change in gene frequency due to selection

= $s.f.(x)$

Equation (3) may be re-written as

$$\frac{\partial \phi(x,t)}{(t/N)} = \frac{\partial^2}{\partial x^2} \left[\frac{x(1-x)\ \phi\ (x,t)}{4}\right] - Ns \frac{\partial}{\partial x} [\phi(x,t)\ f(x)] \quad \ldots(4)$$

From this equation it may be seen that the pattern of change of $\phi_{(x,t)}$ is determined by the product Ns and its time scale is proportional

to N (Robertson, 1960).

Kimura (1957) treated the chance of fixation of a gene by the same approach, in this case using the Kolmogørov backward equation. The solution of this gave chance of fixation as

$$u(p) = \int_0^p e^{-2kDx(1-x) - 2kx}\, dx \Big/ \int_0^1 e^{-2kDx(1-x) - 2kx}\, dx$$

where u(p) = the chance of fixation of A given that its initial frequency was p

and $k = Ns$, $D = 2h - 1$, where hs measures the selective advantage of Aa over aa.

For the additive case, h = 0.5 therefore D = 0.

Therefore $u(p) = \int_0^p e^{-2Nsx}\, dx \Big/ \int_0^1 e^{-2Nsx}\, dx$

$$= \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}} \qquad \ldots(6)$$

With selection for a recessive, h = 0 therefore D = -1

Therefore $u(p) = \int_0^p e^{-2Nsx^2}\, dx \Big/ \int_0^1 e^{-2Nsx^2}\, dx \qquad \ldots(7)$

With selection for a dominant, h = 1 therefore D = 1.

Therefore $u(p) = \int_0^p e^{-2Nsx(2-x)}\, dx \Big/ \int_0^1 e^{-2Nsx(2-x)}\, dx \qquad \ldots(8)$

Robertson (1960) considered these results in detail and referred to u(p) as the "expected limit" to selection. He extended the theory to cover the case of artificial truncation selection for a quantitative trait. Let α = the difference in mean between the two homozygotes AA and aa, in phenotypic standard deviation units. Then s in equations (6), (7) and (8) above may be replaced by $i\alpha$ where i = the selection intensity in phenotypic standard deviation units. The relationship

$$s = i\alpha \qquad \ldots(9)$$

was first suggested by Haldane (1931) and is proved by Kimura (1958) and Griffing (1960). It is an approximation which can be used on the assumption that gene effects are small relatively to the phenotypic standard deviation, i.e. that $\alpha << 1$, that the selection intensity is low, that the phenotypes are normally distributed and that α and σ remain constant throughout the selection process. Latter (1965a) examined this approximation when he applied it to genes of large effect, and found that for 40% of the population selected and $0.5 < \alpha < 1.0$ equation (9) overestimated by 1 - 4% and for 10% of the population selected and $0.5 < \alpha < 1.0$ equation (9) underestimated by 1.6 - 6.2%. Hill (1969a) also considered the errors involved in using the approximation of equation (9) above and found that even for quite large values of α with moderate values of i and small population sizes the errors involved are generally less than 10%. Consequently, equation (9) may be used as a reasonable approximation under a wide range of conditions.

Robertson (1960) considered the time scale of the selection process and introduced the concept of "half-life", this being defined

as the time in generations taken for the mean gene frequency to get half way to the limit. Latter (1966a) also used half-life and, in addition, 95%-life which he defined in the same way. He also defined H50 and H95 as the mean number of generations to 50% and 95% fixation. The absolute value of any system of time measurement is questionable since it may have considerable variance. Kimura and Ohta (1969a) investigated the average number of generations until fixation or loss of a selectively neutral mutant with initial frequency $p = 1/2N_a$; where N_a is the actual population size as opposed to the effective size N. They showed that a fraction $1/2N_a$ of neutral mutants are fixed in approximately 4N generations while the remainder are lost in a few generations. Narain (1970) showed that the standard deviation of the time until fixation of such a mutant is approximately 2N. Kimura and Ohta (1969b) extended their earlier results to confirm this and to show that the mean time until the loss of the neutral mutant is $2(N/N_a)\log_e(2N_a)$ with a standard deviation of approximately $4N/\sqrt{N_a}$.

The equations (6), (7) and (8) provide very simple means for determining the chance of fixation for a single locus and they have been widely applied. Those for the dominance and recessive cases are not quite as useful as for the additive case because they can only be solved numerically. However, this can be quite simply achieved by the application of Simpsons Rule using a high speed computer. For the additive case it has been possible to arrive at at algebraic equations for chance of fixation under population structures other than simple mass selection and this will be shown later.

However, at this point it must again be noted that the partial differential equations method gives only an approximation to the actual Markov chains. To see how good this approximation is, and under what conditions it may reasonably be used, the use of matrix methods must be considered. For any population of size N a transition probability matrix may be set up to study the process of selection. Each element of the matrix, say p_{ij}, gives the probability that the population will be in state j in any generation, given that it was in state i in the previous generation. Where, for example, state i represents the situation in which there are I A alleles and (2N - I) a alleles in the population. I may take values 0,1,2,.....2N so that there are 2N+1 j type states. The matrix therefore has dimensions (2N+1) x (2N+1). The gene frequencies in the initial population can be represented by a vector of dimensions (2N+1), each element of which gives the probability that the population is in any state i. Successive generations of selection are achieved by repeated multiplication of the vector by the matrix. This process is necessarily numerical but it is useful in enabling checks to be made on the validity of the approximate methods. Ewens (1963) used such a system to compare exact values for chance of fixation with the approximation in equation (6). He found that for small s values the diffusion approximation was very close to the true value although in all cases it exceeded it. Hill (1965) extended this comparison to s values approaching unity and found that differences in no case exceeded 8%. Moran (1960) used the transition probability approach, but in this case algebraically, to arrive at

the following equation for chance of fixation

$$u(p) = \frac{1 - e^{-2Np\theta}}{1 - e^{-2N\theta}} \qquad \ldots(10)$$

Where θ cannot be given exactly but lies within limits given by $s(1+s)^{-1} < \theta < s$, where s must be positive.

Therefore, it appears that even for quite fast changes in gene frequency in small populations the results from the partial differential equations approach give a good approximation to chance of fixation for a single locus.

The complications of Linkage and Epistasis

While single locus theory is of great value in increasing our understanding of the operation of selection in finite populations its application has severe limitations. In reality any quantitative trait will be determined by more than one locus and it is to the multilocus system that we must now turn our attention. The introduction of more than one locus immediately involves linkage and possible epistasis which present considerable problems in the extension of relatively simple single locus theory.

Two linked loci

It is again perhaps most useful to consider first the simplest case, that of two loci with additive gene action and no epistasis, for which at least seven parameters must be defined. Consider two loci, each with two alleles, say A and a at the first and B and b at the second locus.

Let

s_1 = selective advantage of A.

s_2 = selective advantage of B.

p = frequency of A in the initial population

q = frequency of B in the initial population.

c = the recombination fraction between A and B.

D = the linkage disequilibrium in the initial population.

N = the effective population size.

These seven parameters uniquely specify the system. However, for convenience the following will also be defined:

f_1 = frequency of the AB gamete in the initial population

f_2 = frequency of the Ab gamete in the initial population

f_3 = frequency of the aB gamete in the initial population

f_4 = frequency of the ab gamete in the initial population

Then $p = f_1 + f_2 \qquad 1 - p = f_3 + f_4$

$q = f_1 + f_3 \qquad 1 - q = f_2 + f_4$ ...(11)

The linkage disequilibrium D is defined as

$$D = f_1 f_4 - f_2 f_3 \qquad \ldots(12)$$

D = O represents linkage equilibrium and in this state the frequency of any gamete is given by the product of the frequencies of the genes it carries. D > O implies an excess of coupling gametes and D < O implies an excess of repulsion gametes. The f_i (i = 1,2,3,4) are uniquely specified by D, p, and q as follows

$f_1 = pq + D$

$f_2 = p(1-q) - D$

$f_3 = q(1-p) - D$ ...(13)

$f_4 = (1-p)(1-q) + D$

D may vary between

D = -0.25, all gametes in repulsion, $f_2 = f_3 = 0.5$

D = +0.25, all gametes in coupling, $f_1 = f_4 = 0.5$

Wright (1933) studied the effects of linkage between two loci in small populations with no selection. These results were extended by Kimura (1963) who developed a general expression for the probability that a chromosome chosen at random carries alleles A and B at time t. This was then found for t = ∞ by the use of a time invariant quantity.

Diffusion methods have been applied to the problem of two linked loci (Kimura, 1955a; Hill and Robertson, 1966; and Ohta and Kimura, 1969) to produce the following equation for the change in gametic frequencies

$$\frac{\partial \phi}{\partial (t/N)} = \frac{1}{4} \sum_{j=1}^{3} \frac{\partial^2}{\partial f_j} [f_j (1-f_j) \phi] - \frac{1}{2} \sum_j \sum_{k} \frac{\partial^2}{\partial f_j \partial f_k} [f_j f_k \phi]$$

$$- \frac{1}{2} Ns_1 \left[\frac{\partial}{\partial f_1} (f_1(1-p) \phi) + \frac{\partial}{\partial f_2} (f_2 (1-p) \phi) - \frac{\partial}{\partial f_3} (f_3 p \phi)\right]$$

$$- \frac{1}{2} Ns_2 \left[\frac{\partial}{\partial f_1} (f_1(1-q)\phi) - \frac{\partial}{\partial f_2} (f_2 q \phi) + \frac{\partial}{\partial f_3} (f_3(1-q)\phi)\right]$$

$$+ Nc \left[\frac{\partial}{\partial f_1} (D\phi) - \frac{\partial}{\partial f_2} (D\phi) - \frac{\partial}{\partial f_3} (D\phi)\right] \quad \ldots (14)$$

where $\phi = \phi (f_1, f_2, f_3, t)$ = the density function of the distribution of gametic frequencies f_i at time t. Therefore, as in the single locus case the time scale of the selection process is proportional to N.

In this case the process is described completely by the initial values of p, q and D and by the parameters Ns_1, Ns_2 and Nc.

Karlin and McGregor (1968) considered in detail the case $s_1 = s_2 = 0$ and analysed Markov chains generated by repeated sampling from the distribution of gamete frequencies in the parent generation. By this method they gave the following equation for chance of fixation of the various gametes

$$u(f_i) = f_i \pm \frac{2Nc}{1 - c + 2Nc} \cdot D \qquad \ldots(15)$$

the sign being negative for i = 1 or 4 and positive for i = 2 or 3.

Ohta (1968) extended Kimura's (1963) probability method which involved recurrence relations from zygote to zygote, to give

$$u(f_i) = f_i \pm \frac{2Nc}{1 + 2Nc} \cdot D \qquad \ldots(16)$$

with sign as above.

Under the model analysed by Karlin and McGregor the generation was considered from one gametic stage to the next and the random process involved was a random union of gametes. However, in the model analysed by Ohta the random process was a random mating of zygotes. The differences arising from these two models are discussed by Watterson (1970) who found that many of the results regarding chance of fixation and rate of fixation are at least qualitatively the same.

Hill and Robertson (1966) considered the rate of breakdown of linkage disequilibrium in the absence of selection, giving the recurrence equation for the mean value of D as

$$D_t = (1-c)(1- 1/2N)D_{t-1}$$

$$= D_o\, e^{-(2Nc + 1)\ t/2N} \text{ approximately} \qquad \ldots(17)$$

Latter (1965b) and Hill and Robertson (1966) used similar computer simulation techniques to consider selection operating under a two locus model. Essentially, they calculated for each gametic type the expected frequency in the next generation, using infinite population theory, and sampled from this frequency distribution using pseudo-random numbers generated by the computer. Both sets of workers used an additive model of gene action with initial linkage equilibrium and considered truncation selection in a quantitative character, letting $s_1 = i\alpha$ as previously defined. Latter restricted his model to the case where $s_1 = s_2$ while Hill and Robertson let $s_2 = i\beta$.

Lewontin and Kojima (1960) developed the necessary equations for selection at two linked loci in an infinite population. These were modified by Hill (1965) to give expected changes in gametic frequencies as

$$\Delta f1 = \frac{i}{2} f1\ (\ \alpha - \beta - \mu\ ) - cD(1 + \frac{i}{2}\ (\ \alpha + \beta - 2\mu))$$

$$\Delta f2 = \frac{i}{2} f2\ (\ \alpha - \mu\ ) + cD(1 + \frac{i}{2}\ (\ \alpha + \beta - 2\mu))$$

$$\Delta f3 = \frac{i}{2} f3\ (\ \beta - \mu\ ) + cD(1 + \frac{i}{2}\ (\ \alpha + \beta - 2\mu))$$

$$\Delta f4 = \frac{i}{2} f4\ (\ -\mu\ ) - cD(1 + \frac{i}{2}\ (\ \alpha + \beta - 2\mu)) \qquad \ldots(18)$$

where $\mu = p\alpha + q\beta$ = the population mean in phenotypic standard deviation units,

and $\Delta p = \frac{i}{2} [\alpha p(1-p) + \beta D]$

$$\Delta q = \frac{i}{2} [\beta q(1-q) + \alpha D] \qquad \ldots(19)$$

If D = O in the initial population, then after one generation of selection

$$D' = -\frac{i^2}{4} f_2 f_3 \alpha\beta = -\Delta p.\Delta q \qquad \ldots(20)$$

This is, in fact, only an approximation. The inclusion of squared terms in gene effects gives

$$D' = \frac{1}{4} (ix - i^2) \alpha\beta p(1-p)q(1-q)$$

where x is the truncation point in standard units (Hill and Robertson, 1966).

Therefore, selection in an infinite population would be expected to generate negative linkage disequilibrium, an effect first pointed out by Nei (1963).

Results of Latter (1965a) and Hill and Robertson (1966) show that with equal effects the ultimate response was reduced by linkage, having the greatest effect when the chance of fixation at both loci, when considered under independent segregation, was in the region of 0.7 to 0.8. However, Hill and Robertson showed that for unequal effects the reduction due to linkage is considerably reduced. They considered their results mainly in terms of the effects of segregation at a second linked locus on the chance of fixation at the first and found that this had no detectable influence unless gene effects at the second locus were greater than half those at the first. That is, only when $\beta > \alpha/2$. Thus for two loci to have mutual influence the condition $2\alpha > \beta > \alpha/2$ must be satisfied. Another important result from this work concerns the linkage disequilibrium which was

built up within lines during selection. This had been expected on the basis of infinite population theory, but even using a multiplicative model of effects, which in an infinite population does not generate disequilibrium (Felsenstein, 1965), considerable linkage disequilibrium was still generated. These results were explained in terms of the effective population size in which gene frequency changes at the locus of smaller effect takes place. Felsenstein (1968) considered this question of interference further, using matrix methods. He concluded that its effect was real and important.

The above studies have shown the complexity of the situation for even the simplest case involving linkage and finite population size. Only a few generalised results have emerged, perhaps the most important is that restriction of population size may drastically effect linkage equilibrium in a way quite unpredictable from infinite population theory.

Many linked loci

Even infinite population theory for several linked loci is complex, although procedures for treating it have been developed, in particular by Geiringer (1944) and Schnell (1961). Work has also been carried out on the infinite population theory by Bennett (1954), who produced formulae for linkage disequilibrium between several loci. However, it is only linkage disequilibrium between pairs of loci which enter into the equations predicting changes of frequency for non-epistatic loci (see below).

For n loci, each with two alleles A_1/a_1, A_2/a_2, ... A_j/a_j ...A_n/a_n,

the linkage disequilibrium between any pair is defined as

$$D_{jk} = f(A_j A_k) - p_j p_k \qquad \ldots(21)$$

where

$f(A_j A_k)$ = the frequency of gamete A_jA_k

p_j = the frequency of allele A_j

p_k = the frequency of allele A_k

Let c_{jk} = the recombination fraction between A_j and A_k. Then, if there is no selection, at any generation t we have

$$D_{jk(t)} = (1 - c_{jk})^t D_{jk(c)} \qquad \ldots(22)$$

For additive loci with effects α_j the change in gene frequency in one generation of selection is given by

$$p_j = \frac{i}{2} (\alpha_j p_j (1-p_j) + \sum_{k \neq j} \alpha_k D_{jk}) \qquad \ldots(23)$$

Lewontin (1965) reviewed the infinite population theory for multiple linked loci and was, in particular, concerned with the effects of epistasis. His approach was to treat the various gametic types as pseudo-alleles and then to use a multi-allelic system. The problems presented by linkage and epistasis in multi-locus systems in finite populations have been tackled by many workers, using a basically descriptive rather than analytical approach. This has been done using computer simulation techniques which will be discussed in detail elsewhere. One of the main workers in the field has been Fraser (1957a) who pioneered many of the methods used. Initially, he made a study of the effect of linkage on rates of advance under selection (Fraser, 1957b) and found a reduction in response with increased linkage. A wide range of problems involving linkage and epistasis for many linked loci have been approached using

this type of simulation and much of this work has been recently reviewed by Fraser and Burnell (1970). Although a great deal of analysis has been carried out in this way there have emerged very few really general conclusions and little has been achieved in terms of increasing our understanding of the way in which linkage and epistasis exert their effects.

Robertson (1969;1970a) has attempted to clarify the position in his theory of limits to artificial selection with many linked loci. His aim was to simplify the situation as far as possible while still being in a position to draw meaningful conclusions. Initially he looked at a set of loci equally spaced along a chromosome, each with two alleles, with equal frequencies and effects at all loci. Initial linkage equilibrium was assumed and an additive model of gene action used. Robertson concerned himself with a "re-parameterization" of the problem and showed that the process of selection can be described by only four parameters, Nih*; Nl, n and q, where

n = number of loci

l = the chromosome length in map units

$h^* = \sigma g^*/\sigma$

σg^*= the genetic standard deviation due to the chromosome

q = the frequency of the favourable allele at each locus.

N and i are as previously defined. In addition he also defined

L_o = the advance at the limit with no recombination

L_f = the advance at the limit with free recombination.

The ratio L_o/L_f was studied under various conditions. In particular he used those values of the parameters most likely to be encountered

in "real life" situations and found that in such cases L_o/L_f is never likely to be less than 0.5 while it may frequently exceed 0.7.

McPhee (1967) selected for high and low sterno-pleural bristle number in Drosophila under two regimes of recombination;

a) Normal recombination

b) Recombination suppressed in chromosomes II and III.

The ratio comparable to Robertson's L_o/L_f was found to be 0.78 for selection upwards and 0.72 for selection downwards.

Robertson (1970) extended his model to cover;

i) inequality of gene effects

and ii) inequality of gene frequencies.

In both cases these tended to increase the ratio L_o/L_f and thus reduce the effect of linkage.

Latter (1969) stressed the need for comparison of simulation theory with observations made on response to selection in living organisms. He compared simulation results obtained under various quite simple models with results obtained from Drosophila. In this way he was able to show that some models were clearly inappropriate.

Throughout this discussion of the theory of selection for linked loci in finite populations the assumption has been made that populations were initially in linkage equilibrium. Under this restriction it has been demonstrated that although linkage can considerably retard progress with selection under an additive model, its influence under what might be termed "realistic parameter sets" may not be of significant magnitude.

CHAPTER II

The importance of the structure of a finite population

This question has been considered mainly from the standpoint of artificial selection where it is of considerable practical importance. In general the aims of artificial selection for a quantitative trait are two-fold:

i) to obtain rapid short term gains, that is to get a nigh immediate response

ii) to utilize as much of the genetic variation available as possible, that is to reach a high limit in the long term.

In practise these two objectives tend to be incompatible, this can be seen by examining the way in which they are affected by the selection intensity.

Consider:

i) the immediate response.

In terms of gene frequency changes the expected response in one generation of selection is given by

$\delta p = i\alpha.f(p)$ where $f(p)$ represents the function of gene frequency dependent upon the type of gene action.

More commonly in quantitative genetics the response to one generation of selection is given as

$R = i\sigma h^2$ (Falconer 1960)

where σ = the phenotypic standard deviation

h^2 = the heritability.

In both cases it can be seen that response is directly proportional to i and is independent of N. Therefore as selection intensity increases so the immediate response increases.

ii) the ultimate limit.

From an earlier section it can be seen that the chance of fixation is a function of the product $Ni\alpha$, increasing as $Ni\alpha$ increases. However N and i may not be independent. Let T be the total number of individuals measured, then N/T is the proportion selected and this is related to i by the equation

$$i = Z/(N/T) = \frac{ZT}{N} \quad \therefore Ni = ZT \qquad \ldots(24)$$

where

Z = the height of the ordinate at the point of truncation.

If T is fixed then Ni is maximized when Z is maximized, which is for 50% selected. This result and its significance was first pointed out by Dempster (1955) and was confirmed by Robertson (1960) for selection for a single locus. With selection for two linked loci Hill and Robertson (1966) showed that the optimum intensity is for slightly more than 50%, approximately 55%, selected. Robertson (1970a) confirms this finding for multiple linked loci. Experimental work using Drosophila has been carried out by Jones, Frankham and Barker (1968) and they confirmed an optimum intensity at around 50% selected.

Smith (1969) and Robertson (1970b) considered the optimum intensity to give maximal gains by a fixed number of generations, say t, for a single additive locus. They showed that the optimum proportion to be selected was a function of t/T but even for t/T=0.5 the proportion selected must be as high as 25%. Therefore unless

a slow rate of initial response can be tolerated, there will be a quite considerable reduction of the limit. Consequently it becomes important to consider whether populations can be structured in any way to improve the long term gains without incurring any reduction in the short term rate of response.

Subdivision of the population

Can subdivision of the population followed by crossing be utilized in any useful way? This question has been examined mainly in terms of utilization of non-additive genetic variation. The general aim being to produce sub-populations which are homozygous and when crossed display useful heterosis. Three basic systems of achieving this have been developed.

a) Pure line selection.

Selection is based on within line performance only

b) Recurrent selection to a tester (Hull,1945).

Selection is carried out in one line only on the basis of its cross performance with a tester strain which may be highly inbred.

c) Reciprocal recurrent selection (Comstock et al,1949)

Selection is based on cross performance and is practised within both populations.

With respect to the utilization of additive genetic variation there has been very little investigation into the role of sub-populations. Robertson (1960) in investigating limits to selection for a single additive locus, showed that if a population of size N is split into x lines each of size N/x and these are selected to

fixation, then crossed and reselected in a single line of size N, that the limit was the same as would have been attained without subdivision. He also showed that on average the subdivision system gave a higher limit when selection was for a recessive, but a lower limit when selection was for a dominant. It is important to note here that the selection intensity was the same in the sub-lines and single lines so that the initial rate of response would be expected to be the same under both systems. However subdivision does slow the overall rate of approach to homozygosity (Robertson 1964).

Pollak (1966) considered natural selection in a subdivided population in which there was migration and he developed an equation giving the chance of fixation of an additive gene by the application of the theory of branching processes. Maruyama (1970) also considered selection operating in natural populations in which there was subdivision with migration. He considered the chance of fixation of an additive gene and extended a method used by Moran (1960) to show that provided the migration does not change the mean gene frequency, the limit was the same as would be achieved under mass selection in a single population of the same total size. This result holds both for inequality of size of sub-populations and for differences in gene frequency between the sub-populations initially.

Madalena (1970) considered a model having five linked loci with equal effects and initial frequencies under initial linkage equilibrium. He used a cyclical structure of inbreeding in sub-lines followed by intercrossing and in some cases practised between line selection. He found that only with intense between line selection was

the response in the cyclical system at any time better than for single mass selection; and in that case the ultimate limit was reduced. Curnow and Baker (1969) examined a theoretical model based on a practical problem concerning selection for yield in maize. They were in particular concerned with the response obtained over a period of 5-10 generations and they showed that this may be higher under a sub-population system. In this case the population was split into sub-populations of relatively small size and selected for a fixed number of generations, then the response in the best line was found to exceed the response in the single large population. The usefulness of this result rests largely on the ability to correctly identify the best line. Madalena (1970) also made a practical study of the cyclical inbreeding and crossing system using Drosophila melanogaster selected for sterno pleural bristle number. He found that in no cases did any of the sub-lining systems exceed the single line in ultimate response, although the situation with respect to rate of response was not as clear.

Bowman and Falconer (1960) working with mice selected for litter size, considered a method of inbreeding and between line selection followed by crossing. They concluded that it was impractical since the population is put through a bottleneck from which it can never subsequently recover.

Even less information seems to be available as to the best way to utilize variation from several distinct populations. James (1966) has considered the question in terms of utilizing separate base populations to form a foundation stock for a selection programme. He compared between line selection with within line

selection and tentatively concluded that there would be a substantial advantage to using stock from several populations rather than simply selecting the best. Jackson and James (1970) carried out a similar practical comparison for wool and body traits in Australian Merino sheep. They found that for a single trait selection from a single population was superior in terms of immediate response but selection from several populations was superior in terms of long term (three or more generations) response. Hill (1971) discussed methods of utilizing breed crosses in animal production and pointed out that although a cross is likely to reach a higher limit than the pure breeds involved, it may take many generations before it is superior to the best pure breed.

In many practical instances effective sub-lining of populations has already been done and populations have been selected until no further response has been obtained. Experimental work; with Drosophila (Clayton and Robertson, 1957; Brown and Bell,1961) showed that populations under continued selection pressure eventually ceased to respond to selection and they found that, in some cases at least, this was due to a loss of additive genetic variance. Falconer and King (1953) and Roberts (1967) crossed lines of mice which had been selected for body weight until they gave no further response, they found on selecting the cross population that renewed response was obtained. Perhaps of more practical importance to animal breeders is the fact that several poultry populations, which were highly selected for egg production, have been plateauing and response has become very low (Dickerson, 1955; Yamada, Bohren and Crittenden, 1958). Robertson (1967) discussed this problem and

mentioned the obvious solution of crossing plateaued lines with other highly selected lines, or alternatively to try to introduce new variation from inferior unselected lines. This latter idea was more fully investigated by Osman and Robertson (1968) both theoretically for a single additive locus and experimentally using Drosophila. Their results favoured selection of the inferior population before crossing and showed no clear advantage in waiting between crossing and reselecting.

The aim of this study has been to consider some of these problems in detail, but first a topic of considerable importance with respect to crossing; either of distinct populations or selected lines, must be considered, and that is linkage disequilibrium.

The importance of linkage disequilibrium

Equation (15) of Chapter I shows the chance of fixation of the various gametes for the case of no selection as given by Karlin and McGregor (1968). Similarly equation (16) gives chance of fixation as derived by Ohta (1968). In both cases the equations show that positive linkage disequilibrium increases the chance of fixation of the coupling gametes and decreases it for the repulsion gametes and vice versa for initial negative linkage disequilibrium. However when there is no selection operating initial linkage disequilibrium has no effect on chance of fixation of individual genes. Hill and Robertson (1966) gave the chance of fixation for a single locus under a two linked additive loci model with only small selective values as

$$u(p) = p + Ns\ p(1-p) + \frac{Ns_2D}{2Nc+1} \qquad \text{approx} \qquad \ldots(25)$$

so that negative initial linkage disequilibrium reduces chance of fixation of individual loci and positive initial linkage disequilibrium increases it, when there is selection. Ohta (1968) considered the effect of initial linkage disequilibrium between two additive loci under selection. She showed that the effect of initial linkage disequilibrium rapidly decreased as Nc increased.

Martin and Cockerham (1960) considered the effect of linkage disequilibrium under a multi locus model. They set up an initial population which was entirely of genotype

A b C d E
a B c D e

with equal effects at all the loci. They found that in this situation there was a very marked effect of linkage as compared with the initial equilibrium situation.

None of these studies give very much insight into the way in which linkage disequilibrium reduces chance of fixation, nor do they give very much indication of the possible conditions under which linkage disequilibrium may be of negligible importance. These problems will be considered in detail later.

The present study

The aim has been to investigate ways of utilizing a sub-line population structure to maximize the selection limit. Two basic situations have been considered with respect to the base population in which selection is to be practised.

a) a single base population in linkage equilibrium

b) two distinct base populations each of which are

themselves in linkage equilibrium.

For the first case the response at the limit for simple mass selection in a single population of size N was compared with the ultimate response obtained using a system of population sub-division and recrossing in a population of total size N.

Similarly for the second case the response at the limit for simple mass selection in a single cross population of size N was compared with the ultimate response obtained using a system of selection in the separate populations followed by crossing again with a total population size throughout of N.

Because these systems of selection and crossing are likely to create linkage disequilibrium initially a study was made of its importance and effect in situations expected to arise from crossing either selected lines or separate populations.

In all the studies made algebraic analysis has been used where possible using existing single locus theory where appropriate, otherwise computer simulation has been used. Before consideration of the results obtained the simulation procedure will be outlined.

CHAPTER III

Simulation Procedure

Many of the problems in theoretical population genetics which have proved to be too complex for any simple algebraic analysis have recently been approached with the help of a high speed computer. A system which has been widely used is "Monte Carlo Simulation" and this was developed by Fraser (1957a). He used the binary nature of an automatic digital computer to represent two alternative alleles at a locus by 1 or 0. Each of these alleles was then assigned a value and the genotypic value of an individual was determined by the values of the alleles it carried and by the genetic model imposed. The phenotypic value of each individual was then given by adding a random environmental component to the individuals genotypic value: the random environmental component was obtained by sampling from a normal distribution with mean zero and known variance. Directional truncation Selection was performed on the basis of phenotypic values by selecting the top P% as parents of the next generation. Recombination and random assortment of gametes gave the next generation of progeny. This system has been used extensively and has been progressively modified to allow a high degree of sophistication in model building.

The main drawback with the basic system originally used by Fraser lies in its expense in terms of computer time. Many of the operations involved for each generation of each replicate of any experiment were in themselves quite lengthy (for example the ordering of individuals prior to selection), while others

required repeated generation of pseudo-random numbers (for example each test for recombination requires one pseudo-random number). Consequently various short-cut techniques have been developed both by Fraser and others, each being appropriate to the type of problem considered. One such system is used in this study for the two locus models, it was developed by Hill and Robertson (1966) to study problems of selection with two linked additive loci in a finite population. The way in which selection was simulated has been outlined earlier, however it should be mentioned that the model used is a "random union of gametes model" as defined by Karlin and McGregor (1968).

The two locus model was used for simulation studies involving only additive loci with a variety of values for the genetic parameters determining gene frequency, gene effects, recombination fraction and initial linkage disequilibrium, and for the operational parameters determining population size and population structure. No attempt has been made to look at all possible comparisons of parameters but instead specific cases have been examined, either as extreme examples to illustrate a particular point, or as typical examples from which more general conclusions can be drawn. As was mentioned earlier the process of selection under a two locus model can be approximately described by the composite parameters Ns_1, Ns_2, Nc (the accuracy of this approximation is discussed by Hill and Robertson (1970)) together with p, q and D, this fact was utilized in keeping N low at all times (in general N = 8 was used though some runs with N = 16 were also done as checks) and varying s_1, s_2 and c.

With model systems involving more than two loci a different approach has been used. The computations made in this study were carried out mainly on the Edinburgh Regional Computing Centre's ICL system 4-75 Machine using IMP as the compiler language. Because of the complexity of this system any attempt at using the binary nature of the computer raised more problems than it solved. Consequently a rather different system was adopted, each parent in the population at any time was represented by a 2xn array, (where n was the number of loci considered) of ones and zeros from which a genotypic value could be calculated as with the binary method. Instead of directly simulating selection as Fraser (1957a) did, a short cut method developed by Robertson (1970a) was used. He considered the probability that any individual parent would contribute a gamete to the N parents which would be selected in the next generation, this method is strictly only applicable to additive loci with no epistasis. One way to approach the problem is to consider a hypothetical progeny generation with phenotypic variance σ_p^2 and mean μ, -which is also the mean of the parent population. Let all values be given as deviations from the progeny mean in phenotypic standard deviation units, so that the proportion selected, P, is given by

$$P = \int_x^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \qquad \ldots(26)$$

where x is the point of truncation.

Let any individual parent, I, have a genotypic value $2g_i$, then if it was mated at random with the rest of the population it

would produce a population of progeny with mean g_i and variance

$\sigma^2_e + \frac{\sigma^2_a}{4}$ (Falconer 1960),

where σ^2_e = the environmental component of variance plus variance due to other loci

σ^2_a = the additive genetic variance due to the loci considered.

The probability that any individual I will contribute to the next parent generation is given by the proportion of its progeny which would be expected to be selected as parents, let this be Wi. To find Wi the truncation point must be converted to a deviation from the I progeny mean in the appropriate phenotypic standard deviation units.

$$\therefore \text{ let } x^1 = (x - g_i) \frac{\sigma_P}{\sqrt{\sigma^2_e + \sigma^2_a}}$$

$$\text{then } Wi = \int_{x^1}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}}$$

$$= \int_{\frac{(x-g_i)\sigma_P}{\sqrt{\sigma^2_e + \sigma^2_a}}}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \, dy \qquad \ldots(27)$$

if σ^2_e is large compared with σ^2_a then $x^1 \approx (x - g_i)$

and

$$W \approx \int_{(x-g_i)}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \, dy \qquad \ldots(28)$$

This approximation was used in all cases with

$$g_i = \sum_{j=1}^{n} \alpha_j \; p_{ij}$$

where

α_j = the effect of the favourable allele at the jth locus

p_{ij} = the frequency of the favourable allele at the jth locus in the ith individual

Each individual was given a "Fitness Value", F_i, proportional to Wi such that

$$F_i = \frac{Wi}{\sum_{i=1}^{N} Wi} \qquad \therefore \sum_{i=1}^{N} F_i = 1$$

The next generation was formed by sampling from the distribution of fitness values, this selected a parent from which a gamete was extracted by a process involving recombination. In this way 2N gametes were formed and combined at random to give the new parent generation. In the simulation done here the truncation point x was taken for all cases as 0.25 standard units, this corresponds to a proportion selected of approximately 40%. In most cases 10 parents were selected each generation so that P = 10/25, the selection intensity which this gives is obtained from tables (Becker 1967) as i = 0.936. With N small, effects large and few loci the relationship between P and x is no longer strictly true and errors involved in using this system have been considered by comparison with a 'true' truncation selection system as used by Fraser (1957a). Results are given below for N = 10, equal effects and frequencies for all loci and equal recombination fractions between all adjacent pairs

Niα	Nc	q	Mean chance of fix^n 'True'Sel^n	Approx. Sel^n	Diff	t
5	0	0.2	0.5175	0.5250	-0.0075	0.31
		0.3	0.6716	0.6600	+0.0116	0.47
		0.5	0.8600	0.8300	+0.0300	1.41
5	0.3125	0.2	0.6300	0.5960	+0.0340	1.17
		0.3	0.7975	0.7350	+0.0625	2.51
		0.5	0.9550	0.9200	+0.0350	2.57
2.5	0	0.3	0.4800	0.4850	-0.0050	0.13
7.5	0	0.3	0.8250	0.800	+0.0250	0.73

This shows that the approximations used tend to reduce the estimate of chance of fixation although the greatest reduction found was less than 10%. Since most of the work using this approximation has been concerned with comparisons of chance of fixation under different selection systems, the errors involved are not considered to be of any great importance. Recombination can be achieved quite simply by performing a "random walk" along the chromosome pair, at each point between a pair of loci the decision whether or not to cross over on to the other chromosome is made by comparing a pseudo-random number with the recombination fraction c_{ij}.

For closely linked loci, say $c < 0.1$, it can be assumed that there will not be more than one crossover between any adjacent pair of loci. Under these conditions a short-cut method described by Robertson (1970a) was used. Recombination was specified in map units by the equation

$$r_{ij} = \frac{1}{2} \log_e (1 - 2 c_{ij}) \qquad \ldots(29)$$

where

r_{ij} = the map distance between the ith and jth locus.

The probability that no recombinational events will occur in any small distance r_{ij} is given by $e^{-r_{ij}}$, so that one pseudo-random number, x, gives the distance over which no recombination will occur as y, where $x = e^{-y}$, $\therefore y = -\log_e(x)$. y also gives the point of crossover and after changing chromosomes a new y is generated and so on until the end is reached. This system was used for all cases where c was <0.1.

The sections of my programmes performing both the short-cut recombination and selection methods were given to me by J.W.James although the original systems are due to Robertson.

In all the simulation programmes each replicate was selected until either all loci had become fixed or for 6.25 N generations (this time being based on the response obtained for a single locus (Hill & Robertson, 1966)) whichever occurred first. The number of replications used varied depending upon the accuracy required for any particular investigation. For the multi locus programmes 100 replicates was the usual figure used, while for some of the investigations under the two locus model 1600 replications were made.

CHAPTER IV

The Effect of Initial Linkage Disequilibrium

(a) Introduction

Any cross between two genetically distinct populations may generate linkage disequilibrium and for this reason the effect of initial linkage disequilibrium on response to selection may be of considerable importance.

As an example consider two loci with alleles A/a and B/b segregating in two populations 1 and 2.

Let

Frequency of allele A in population 1 = p_1

" " A " 2 = p_2

" " B " 1 = q_1

" " B " 2 = q_2

Suppose each population is itself in linkage equilibrium and that the cross is made by combining equal numbers of individuals from each population, then D is given by

$$D = \frac{1}{4}(p_1 - p_2)(q_1 - q_2) \qquad \ldots(30)$$

$\therefore$ if $p_1 = p_2$ or $q_1 = q_2$ then $D = 0$

If $p_1 > p_2$ and $q_1 > q_2$ or $p_2 > p_1$ and $q_2 > q_1$, then $D > 0$

if $p_1 > p_2$ and $q_1 < q_2$ or $p_2 > p_1$ and $q_2 < q_1$, then $D < 0$

If the two populations are lines selected to fixation p_1, p_2, q_1 and q_2 can only take values of zero or one and therefore D can only take values of -0.25, 0 or +0.25. For D=0 at least one locus must be fixed in the cross, so that consideration of both loci segregating only concerns cases where disequilibrium is expected.

For this extreme case there are two possible situations, viz:

i) only AB and ab gametes segregating with D = 0.25

ii) only Ab and aB gametes segregating with D = -0.25

In the former case A and B may be fixed simultaneously and the main effect of the disequilibrium will be to enhance chance of fixation, the amount by which this is done depending upon the size of effects. However in the latter case, in the absence of recombination, only A or B can be fixed and the disequilibrium will be expected to have a considerable effect on chance of fixation of the loci, in particular if they are of equal effect when the maximum for each becomes 0.5. For this reason it is the case of negative initial disequilibrium and its effects which has been considered in detail.

If a cross population is to be constructed from two separate populations this may be done in one of two ways

i) by allowing all individuals to mate at random, this will be termed a "Random Cross".

ii) by specifically mating individuals from one population with individuals chosen at random from the other population, thereby making an F_1 cross.

If there is no recombination and no selection among F_1 individuals then the F_2 generation is equivalent to the random cross. The gamete frequencies in the F_2 are given by

$$f_1' = f_1 - 2cD$$

$$f_2' = f_2 + 2cD$$

$$f_3' = f_3 + 2cD$$

$$f_4' = f_4 - 2cD \quad \text{if there is no selection.}$$

The linkage disequilibrium in the F_2 is given by

$$D_{F_2} = (1-2c)D \qquad \ldots(31)$$

The effect of this generation can be seen by considering the chance of fixation for the no selection case.

Let

$u(f_1)_R$ = the chance of fixation of AB from a random cross

$u(f_1)_F$ = the chance of fixation of AB from an F_1 cross

then from (16)

$$u(f_1)_R = f_1 - \frac{2Nc}{1+2Nc} D$$

$$u(f_1)_F = f_1' - \frac{2Nc}{1+2Nc} D_{F_2} = f_1 - \frac{(2c + 2Nc)}{1 + 2Nc} D$$

$$\therefore \frac{u(f_1)_F - f_1}{u(f_1)_R - f_1} = 1 + \frac{1}{N} \qquad \ldots(32)$$

i.e. the ratio of the response under the F_1 cross to that under the random cross approaches 1 as N becomes large.

The effect of an unselected F_1 generation on an otherwise selected population can only be seen by simulation. This has been done for an extreme case where $p_1 = q_2 = 1$, $p_2 = q_1 = 0$ and $D = \frac{1}{4}$. Figure 4.1 shows results for $N = 8$, $Ni\alpha = 4$, $Ni\beta = 4.2$. $u(f_1)_R$ and $u(f_1)_F$ show a marked linearity with $\frac{Nc}{2(1+2\,Nc)}$, the F_1 in general giving a higher chance of fixation. These results suggest that in such an extreme situation at least, the important factor with or without selection is the probability of forming the AB gamete which appears to be a function of $\frac{Nc}{2(1+2Nc)}$ in both cases.

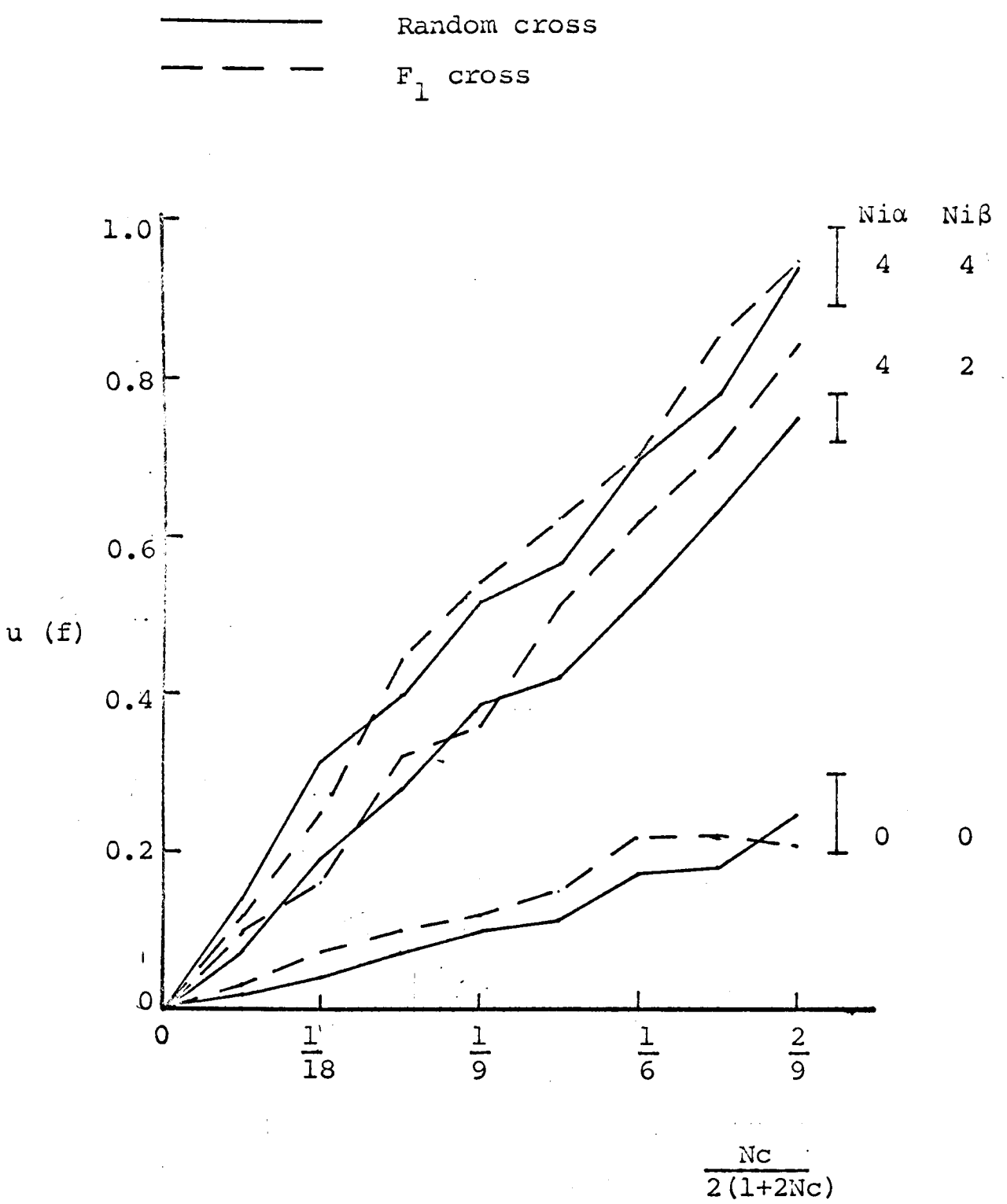


FIGURE 4.1. The effect of an F_1 cross on chance of fixation of the AB gamete for N = 8, n = 2, p = q = 0.5. Do = 0.25. Typical ranges of length four standard errors are also shown.

The importance of not practising selection in the F_1 generation will be considered elsewhere but clearly if disequilibrium is extreme and effects equal no selection can operate anyway. On the other hand if the disequilibrium is less extreme such that a few AB gametes are present, an unselected generation may be highly disadvantageous in allowing chance loss of such a valuable gamete.

Consider next a more general multi locus situation. With n loci, let

P_{1i} = the frequency of the favourable allele at the ith locus in population 1.

P_{2j} = the frequency of the favourable allele at the ith locus in population 2.

D_{ij} = the disequilibrium between the ith and jth loci.

c_{ij} = the recombination fraction between the ith and jth loci.

Then for the cross situation described above

$$D_{ij} = \frac{1}{4} (P_{1i} - P_{2i}) (P_{1j} - P_{2j}) \qquad \ldots(33)$$

Similarly

$$D_{jk} = \frac{1}{4} (P_{1j} - P_{2j}) (P_{1k} - P_{2k})$$

now if D_{ij} and D_{jk} are both negative then

$D_{ik} = \frac{1}{4} (P_{1i} - P_{2i}) (P_{1k} - P_{2k})$ must be positive i.e. for every two pairs of loci in negative linkage disequilibrium at least one pair must be in positive disequilibrium. This will be

discussed in detail elsewhere.

If an F_1 cross is made and is unselected then

$$D_{ij}\ F_2 = (1 - 2\ c_{ij})\ D_{ij} \qquad \ldots(34)$$

So the magnitude of both the negative and positive disequilibria will be reduced, this will increase chance of fixation for gametes containing pairs initially in negative disequilibrium but will reduce it for those containing pairs initially in positive disequilibrium. Therefore the effect of such a generation will depend upon the relative importance of the constituent pairs. Again an unselected generation may allow loss of important rare gametes which carry several favourable alleles. Simulation results show that even with maximum negative disequilibrium and intermediate recombination fractions, the F_1 cross may give a significantly lower response than the random cross, but all differences are small.

In the simulation studies made during the course of this work the F_1 cross has always been made under the two locus model since this in general gives the higher response. However since this appears no longer to be true under the multi locus model, a simpler random cross has in many cases been used.

In order to examine the effects of initial negative linkage disequilibrium in a finite population a detailed study has been made under a restricted two locus model. Following this a limited study has been made for the multi-locus case to see to what extent the conclusions from the two locus study are generally applicable.

(b) Results under the two locus model.

The effect of initial linkage disequilibrium was studied by comparing the chance of fixation and response in the mean, with a population with the same loci at the same frequencies but initially in linkage equilibrium.

A special case giving maximum linkage disequilibrium has been studied where $p_2 = q_1 = 0$ so $D = -\frac{1}{4} p_1 q_2$ for brevity let $p_1 = 2p$, $q_2 = 2q$ then $D = -pq$ and p and q are the frequencies in the cross.

The response in the mean was measured in two ways:

i) The mean proportional response (M.P.R.)

This is the observed change in the mean at the limit as a proportion of that change which would have been acnieved if all favourable genes had been fixed (Latter 1966a). So for two loci A and B

$$M.P.R. = \frac{(u(p)\alpha + u(q)\beta) - (p\alpha + q\beta)}{(\alpha + \beta) - (p\alpha + q\beta)} \qquad \ldots(35)$$

ii) The mean absolute response (M.A.R.)

This is the observed change in the mean at the limit as a proportion of the overall range. So

$$M.A.R. = \frac{(u(p)\alpha + u(q)\beta) - (p\alpha + q\beta)}{\alpha + \beta}$$

$$= M.P.R. \left(1 - \frac{p\alpha + q\beta}{\alpha + \beta}\right) \qquad \ldots(36)$$

This may give a more useful measure in practical terms.

Clearly an important factor influencing the effect of initial linkage disequilibrium is the degree of recombination, therefore results will be presented for:

i) Free Recombination

ii) No Recombination

iii) Intermediate Recombination

i) Free Recombination

When c = 0.5 the loci are unlinked and if an unselected F_1 cross is made then

$$D_{F_2} = D\ (1-2c) = 0$$

i.e. there will be no effect of the initial disequilibrium since it disappears in the first generation.

ii) No Recombination

When c = 0 the loci are completely linked and the F_1 cross can have no effect in changing the disequilibrium. Under the model considered here the AB gamete can never appear in the disequilibrium population and the importance of this will be discussed for (a) Equal effects and (b) Unequal effects.

(a) Equal Effects.

$\alpha = \beta$. In this case chance of fixation in the disequilibrium population may be treated algebraically using equation (6). Let $u(p)_D$, $u(q)_D$, ~~$u(q)_D$~~ and $u(f_i)_D$ represent the chance of fixation of the genes and gametes in the disequilibrium population and similarly $u(p)_E$, $u(q)_E$ and $u(f_i)_E$ t hose in the equilibrium population, then

$$u(f_2)_D + u(f_3)_D = u(p)_D + u(q)_D = \frac{1-e^{-2Ni\alpha(p+q)}}{1-e^{-2Ni\alpha}}$$

and

$$u(f_2)_D = u(p)_D = \frac{p}{p+q} \quad \frac{1-e^{-2Ni\alpha(p+q)}}{1-e^{-2Ni\alpha}}$$

$$u(f_3)_D = u(q)_D = \frac{q}{p+q} \quad \frac{1-e^{-2Ni\alpha(p+q)}}{1-e^{-2Ni\alpha}} \qquad \ldots(37)$$

In the equilibrium population chance of fixation can only be obtained by simulation but since $\alpha = \beta$

$$u(f_2)_E = \frac{p}{p+q} \quad (u(f_2)_E + u(f_3)_E)$$

and $$u(f_3)_E = \frac{q}{p+q} \quad (u(f_2)_E + u(f_3)_E)$$

$$\therefore \quad u(p)_E = (u(f_1)_E + \frac{p}{p+q} \quad (u(f_2)_E + u(f_3)_E)$$

$$\therefore \quad u(p)_E - u(p)_D = u(f_1)_E + \frac{p}{p+q} (u(f_2)_E - u(f_2)_D + u(f_3)_E - u(f_3)_D)$$

$\ldots(38)$

Now $u(f_2)_D > u(f_2)_E$ and $u(f_3)_D > u(f_3)_E$ since f_2 and f_3 are higher in the disequilibrium population and also the Ab and aB gametes face no selective competition as they do in the equilibrium population. Therefore the term $\frac{p}{p+q} (u(f_2)_E - u(f_2)_D + u(f_3)_E - u(f_3)_D)$ will be negative in sign and since

$$u(q)_E - u(q)_D = u(f_1) + \frac{q}{p+q} (u(f_2)_E - u(f_2)_D + u(f_3)_E - u(f_3)_D)$$

it is apparent that the effect of the disequilibrium will be greatest for the lower frequency allele.

If pq is low and effects are not large $u(f_1)_E$ will be small and so will be differences in chance of fixation in Ab and aB, therefore only small difference would be expected for either locus.

If pq is high and effects are quite large then $u(f_1)_E$ will approach 1. $u(f_2)_E$ ~~and $u(f_2)_E$~~ and $u(f_3)_E$ will approach zero and $u(f_2)_D + u(f_{3D})$ will approach 1 in which case $u(p)_E - u(p)_D$ will

approach $\frac{q}{p+q}$. Figure 4.2 shows simulation results with $u(p)_E$, $u(q)_E$, $u(p)_D$ and $u(q)_D$ plotted against q. For p=q = 0.05 the difference in chance of fixation is non-significant being of the order of 0.04, while for p = q = 0.5 the difference is of the order of 0.50. Figure 4.3 gives the difference in M.A.R. plotted against pq and this shows a marked curvilinear relationship, this might be expected since pq is clearly of importance in determining the difference in chance of fixation between the disequilibrium and equilibrium populations. To examine this point further two alternative situations were considered, viz:

i) pq held constant with p + q varied.

ii) p+q held constant with pq varied.

i) pq held constant.

Figure 4.4 shows simulation and theoretical (for the disequilibrium case) results for pq = 0.0625 with p + q varying between 0.500 and 0.625, (p increasing and q decreasing as (p+q) increased). Although obscured somewhat by sampling in the equilibrium case it appears that both populations behave in a very similar manner with changes in (p+q) resulting in the observation that the difference in M.A.R. remains constant for all p+q.

ii) p+q held constant.

Figure 4.5 shows results for p+q = 0.6 with pq varying between 0.05 and 0.09. As p increases q decreases by as much and so $u(p)_D$ increases and $u(q)_D$ decreases by the same amount, this is apparent from (37). However in the equilibrium population $u(f_1)_E$ increases with increased pq such that $u(p)_E$ increases by more than

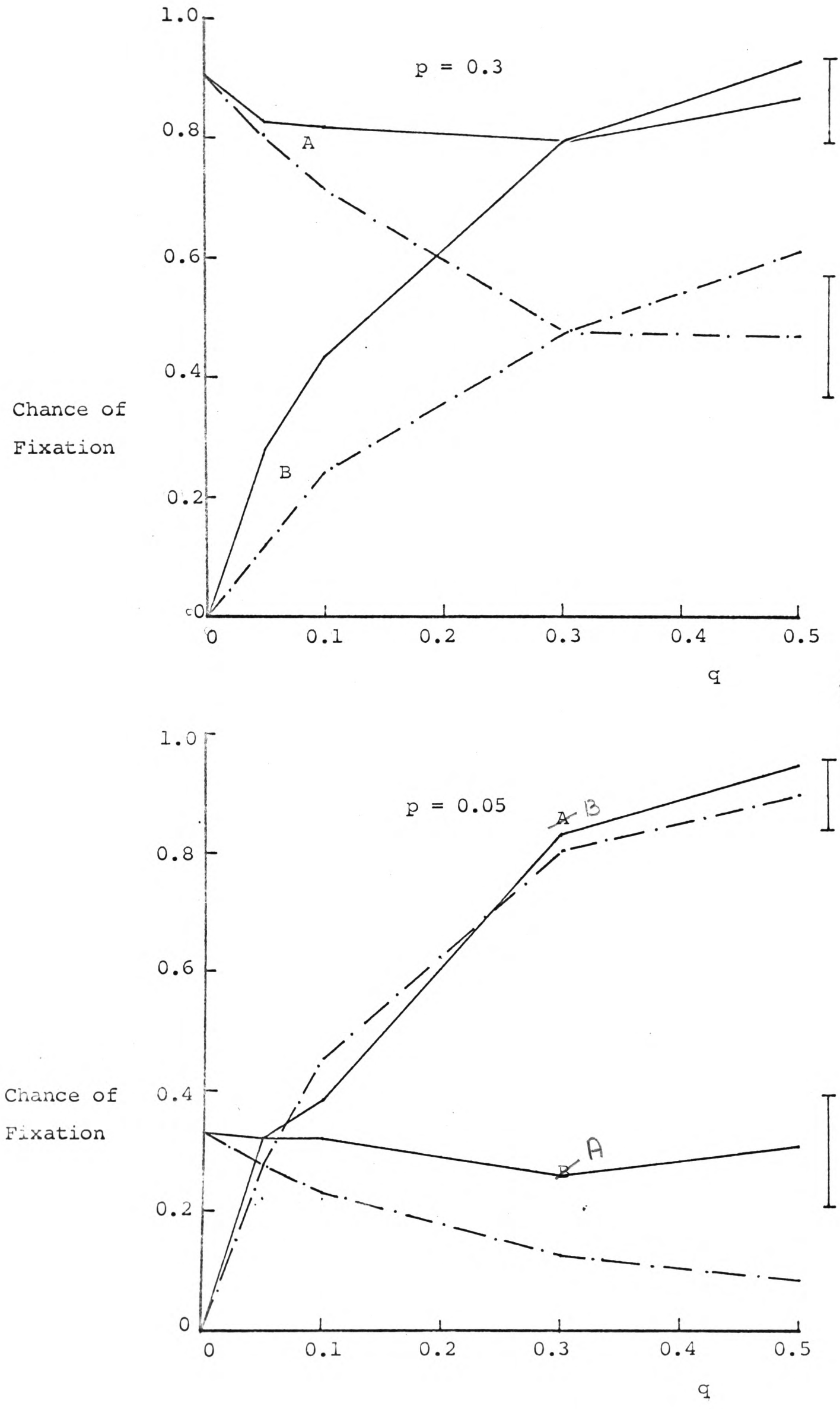


FIGURE 4.2. The effect of initial linkage disequilibrium on chance of fixation of favourable alleles for N=8, n=2, Niα = Niβ =4. Nc = 0 and various values of p and q. Typical ranges of length four standard errors are also shown.

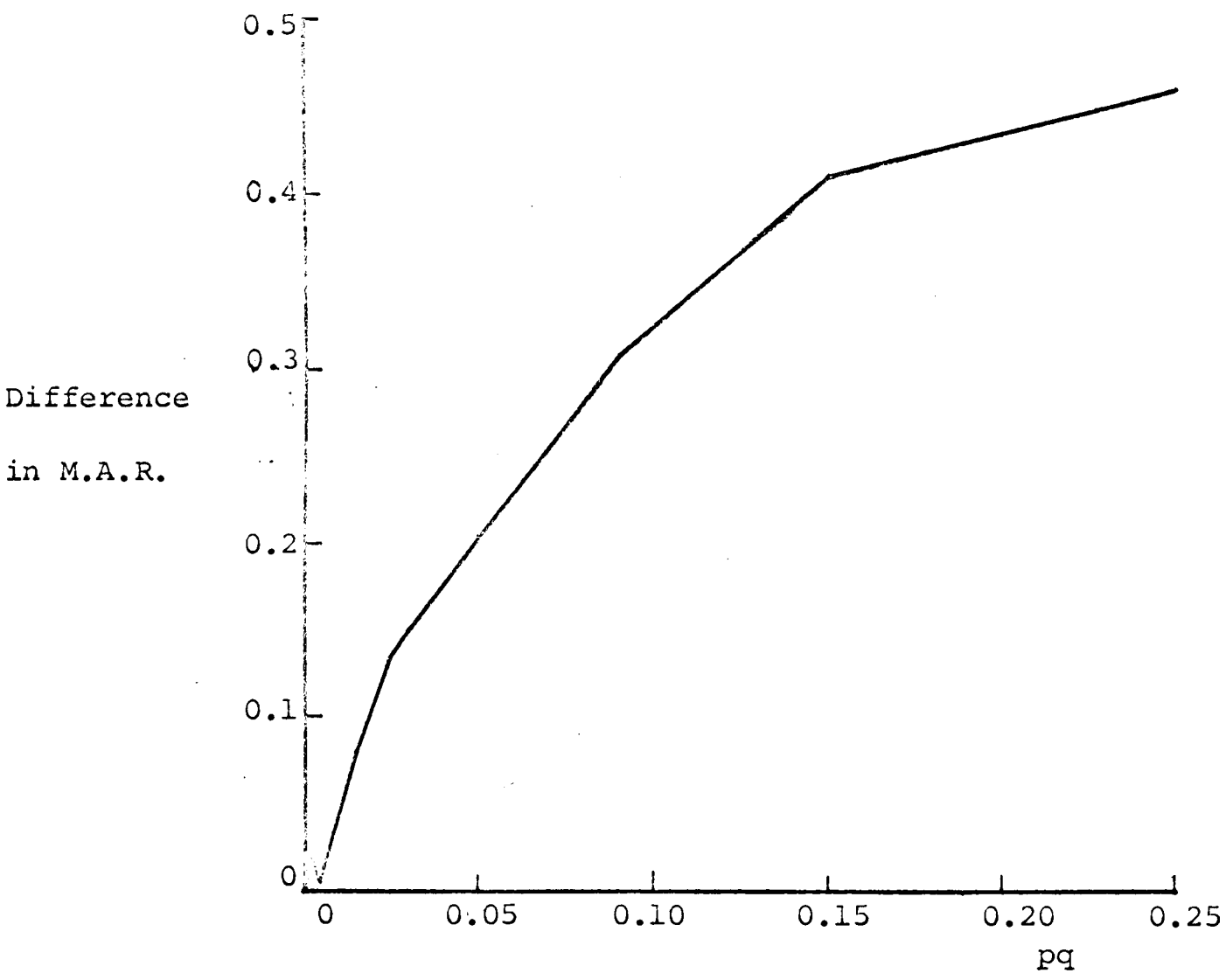


FIGURE 4.3. The relationship between the difference in mean absolute response and pq for N = 8, n = 2. Niα = Niβ = 4, Nc = 0

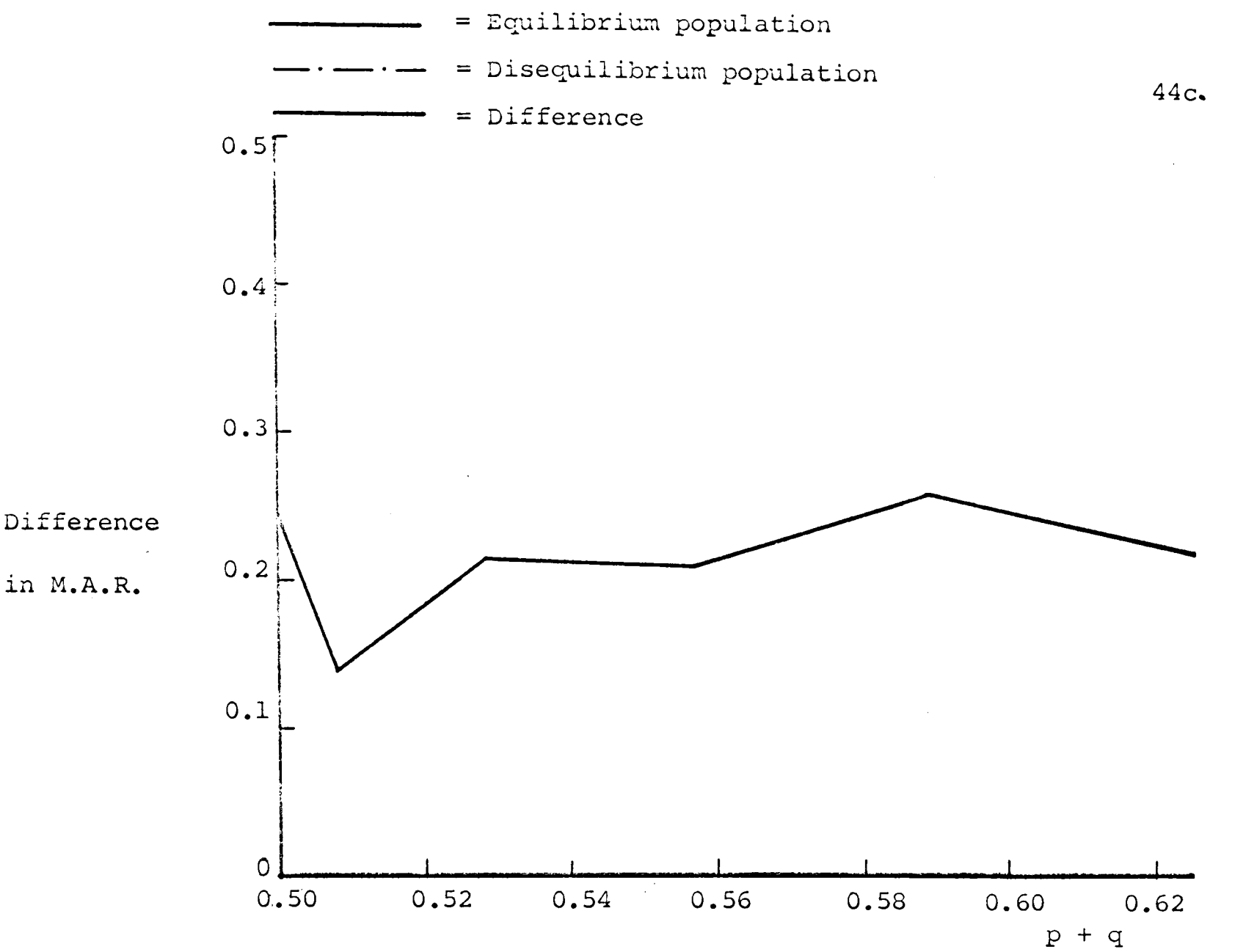
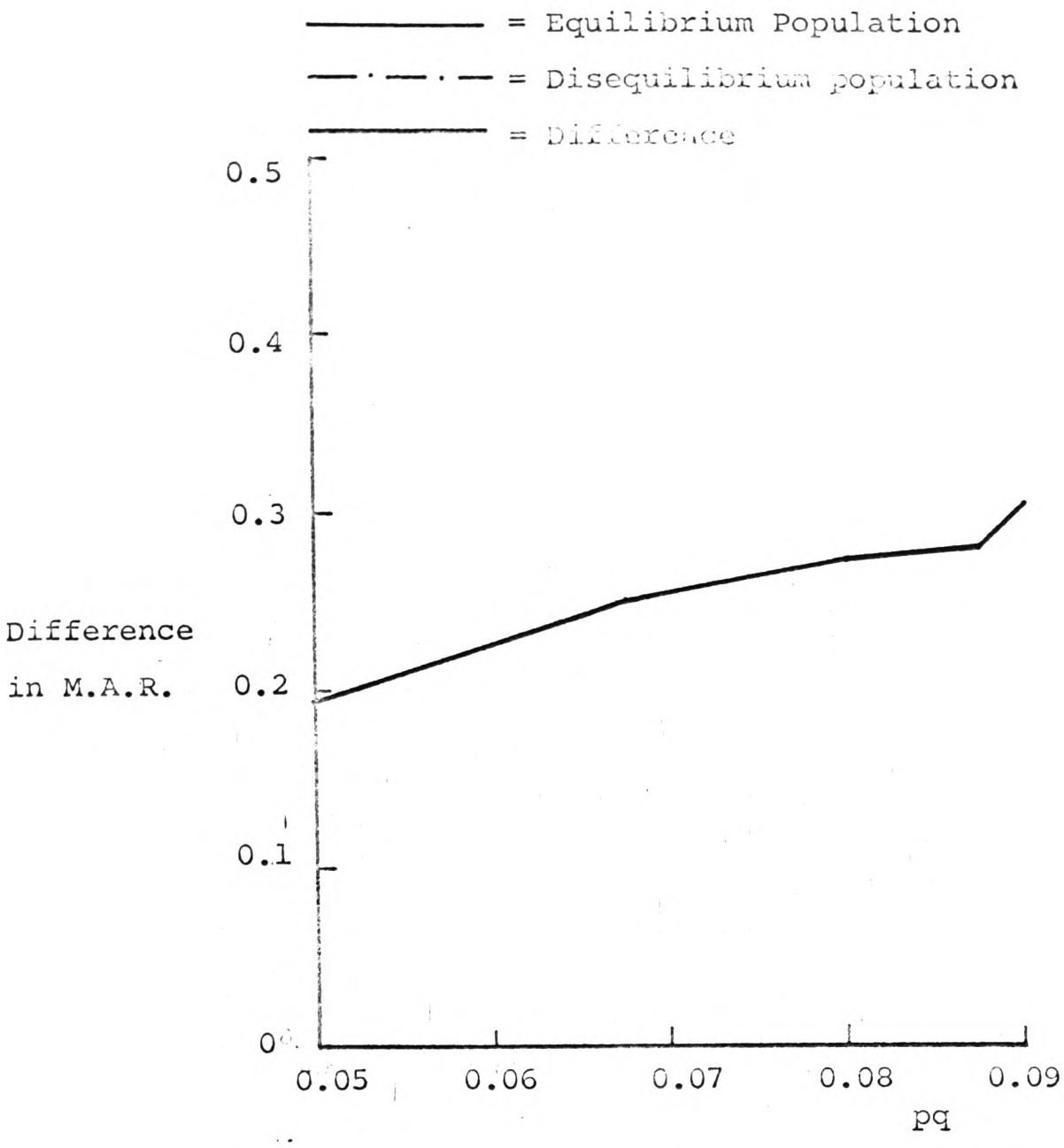


Figure 4.4 The influence of p+q on the effect of initial linkage disequilibrium for N=8, n=2, Niα = Niβ = 4. Nc = 0, pq = 0.0625. Typical ranges of length four standard errors are also also shown.



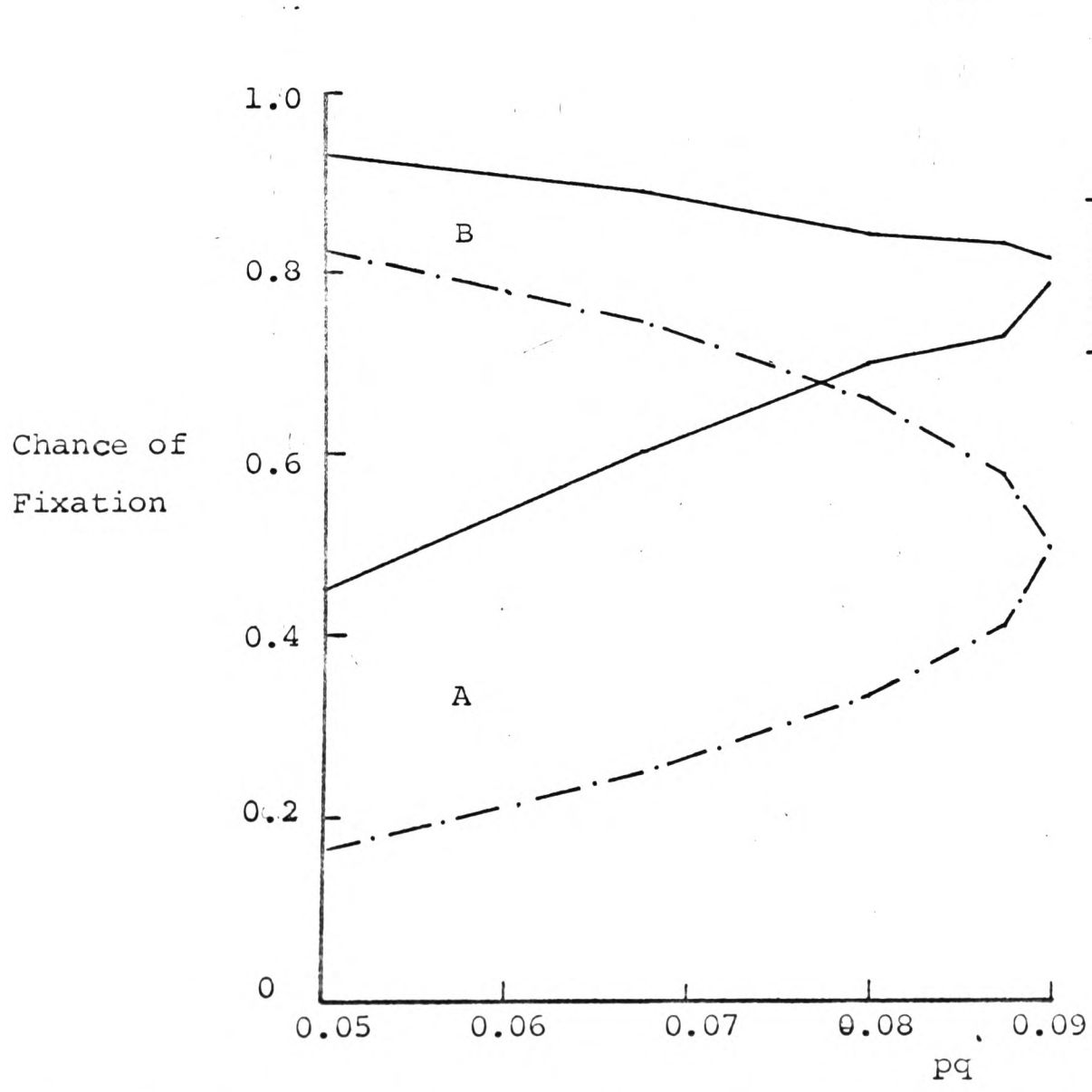


FIGURE 4.5 The influence of pq on the effect of initial linkage disequilibrium for N=8, n=2, Niα = Niβ =4, Nc=0, p+q=0.6. Typical ranges of length four standard errors are also shown.

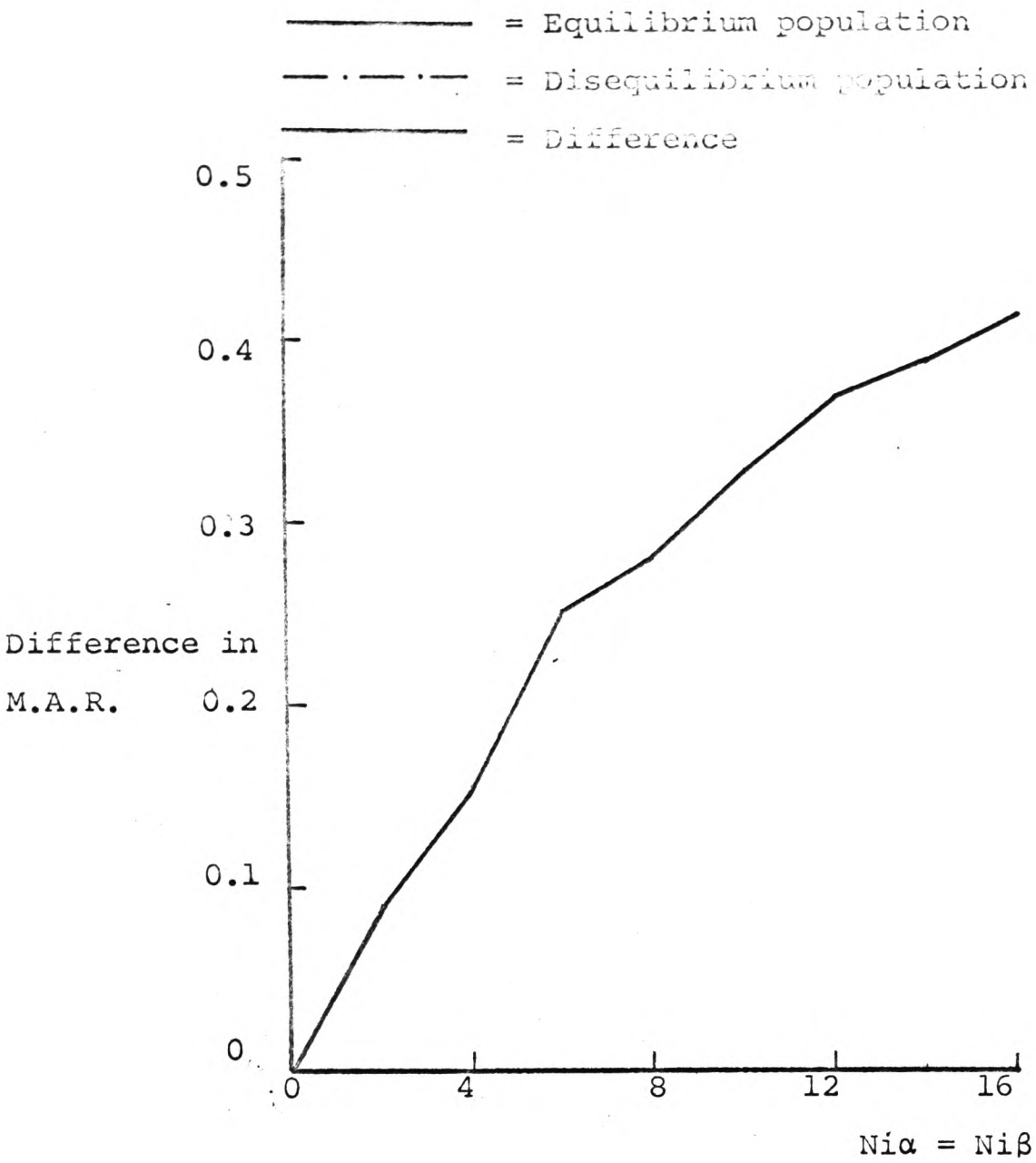
$u(q)_E$ decreases. Therefore the difference in M.A.R. increases with pq as expected.

These studies have shown that althougn the difference in response between the disequilibrium and equilibrium populations may not be entirely described by the composite parameter pq, it does account for most of the variation. Before leaving the equal effects case consider the importance of the absolute magnitude of gene effects. Clearly if $p = q = 0.5$ the magnitude of the effects can make no difference to the disequilibrium population as chance of fixation for all $\alpha = \beta$ is constant at 0.5. Therefore the difference in M.A.R. will increase as the response on the equilibrium increases reaching a maximum value of 0.5 for large $N_i\alpha$. If p and q are less than 0.5 then $u(p)_D$ and $u(q)_D$ will increase towards $\frac{p}{p+q}$ and $\frac{q}{p+q}$ respectively as $N_i\alpha$ increases.

Results for $p = q = 0.2$ are given in Figure 4.6 showing $u(p)_D$ and $u(q)_D$ approaching 0.5 for large $N_i\alpha$ as $u(p)_E$ and $u(q)_E$ approach 1. The difference in M.A.R. increases from zero towards a maximum of 0.5 for large $N_i\alpha$.

(b) Unequal effects.

If gene effects are not equal then the three gametes segregating in the disequilibrium population are all of different selective value and consequently single locus theory can no longer be applied. However the extreme case of $p = q = 0.5$ gives only two gametes segregating, then



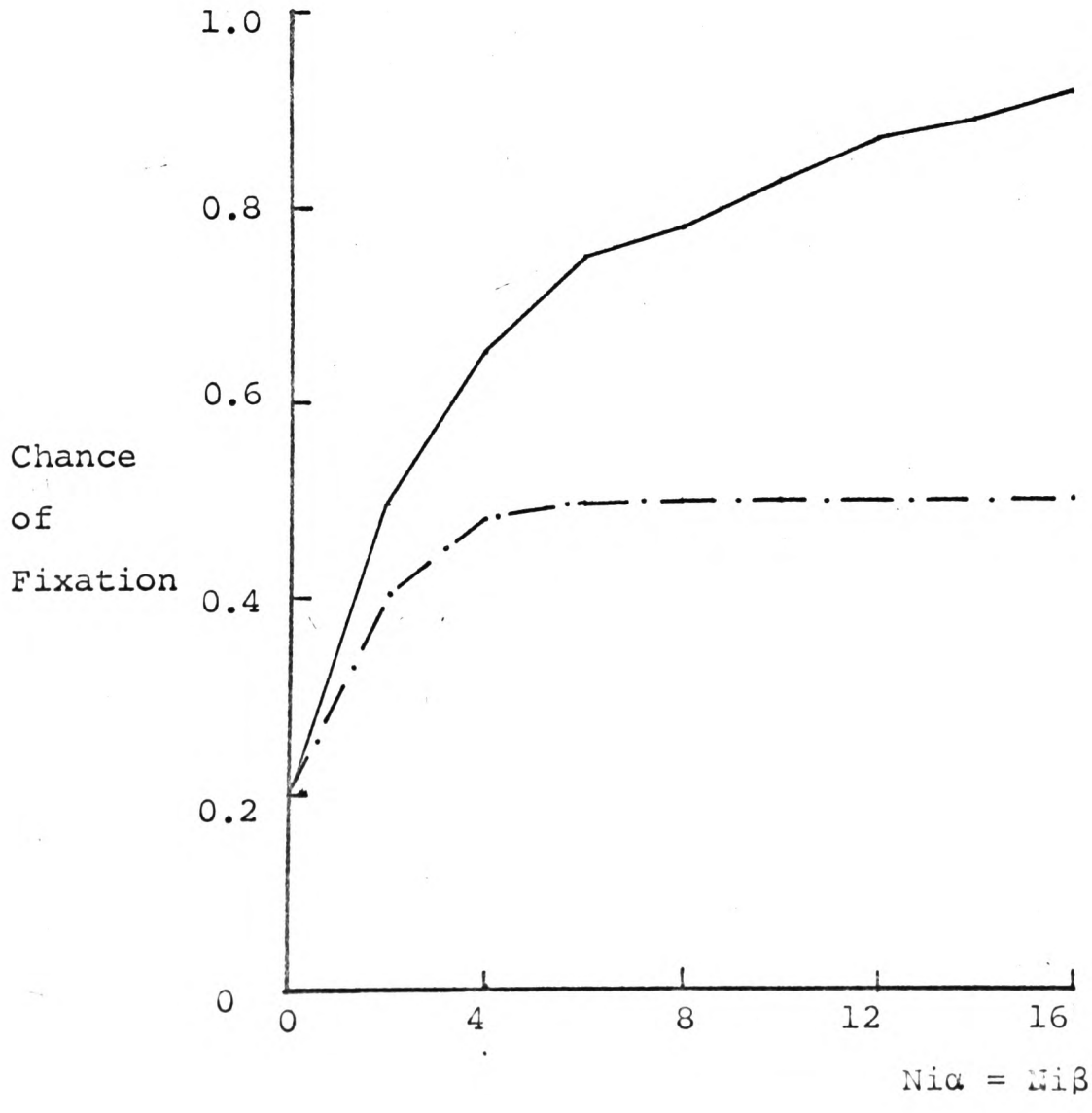


FIGURE 4.6. The influence of the absolute magnitude of gene effects on the importance of initial linkage disequilibrium for N=16, n=2, Nc=0, p=q=0.2. Typical range of length four standard errors is also shown.

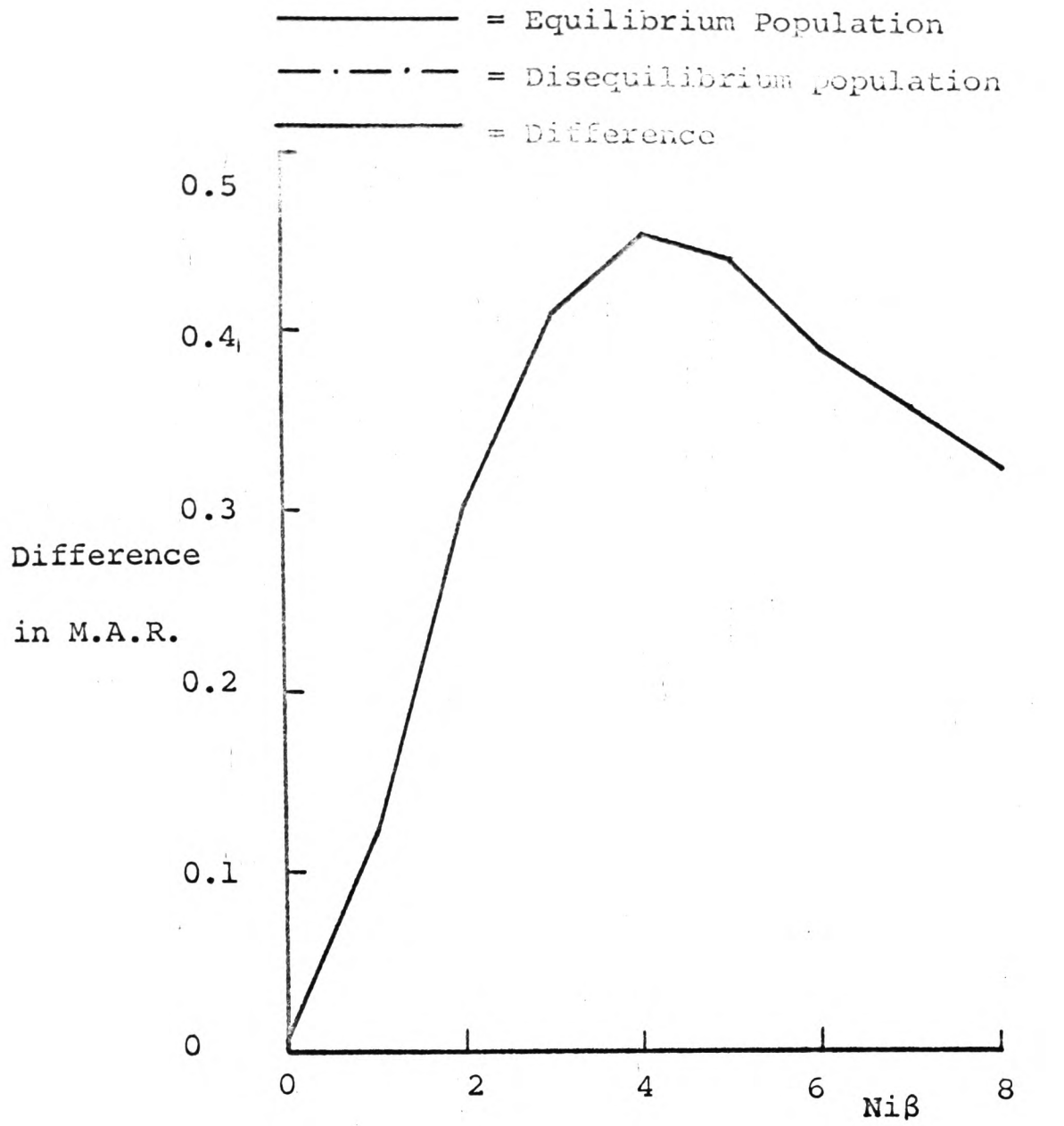
$$u(f_2)_D = u(p)_D = \frac{1-e^{-Ni(\alpha-\beta)}}{1-e^{-2Ni(\alpha-\beta)}} \qquad \ldots(39)$$

$$u(f_3)_D = u(q)_D = 1 - u(p)_D = \frac{1-e^{-Ni(\beta-\alpha)}}{1-e^{-2Ni(\beta-\alpha)}}$$

Let $\alpha > \beta$ then the largest gamete present in the disequilibrium population is Ab and that in the equilibrium population is AB. The ratio of the value of AB to that of Ab is given by $\frac{\alpha+\beta}{\alpha} = 1+\beta/\alpha$. If $\alpha = \beta$ this ratio is 2, otherwise it approaches 1 as α increases or β decreases in value, in which case the disequilibrium will be of negligible importance in terms of over all response.

For a fixed value of α consider the effect of varying β when $p = q = 0.5$. As β increases from zero to some value considerably greater than α so $u(q)_D$ increases from zero to one and $u(p)_D$ decreases from one to zero, if α is reasonably large. In the equilibrium population chance of fixation can only be obtained by simulation and figure 4.7 gives results for $Ni\alpha = 4$ with $Ni\beta$ varying between zero and 8. In this case $u(p)_E$ remains high for all β and $u(q)_E$ increases from 0.5 to 1. It is apparent from this figure that disequilibrium has most effect on the smaller allele. Since the M.A.R. is mainly determined by the larger locus the difference tends to zero as $\beta \to o$ and as $\beta \to \infty$ and is maximized as expected for $\alpha = \beta$.

Consider next the effect of gene frequency changes for the case $\alpha \neq \beta$. Figure 4.8 gives simulation results for a range of p and q values and this shows that for $\alpha > \beta$, the disequilibrium has most effect in reducing the chance of fixation of the smaller



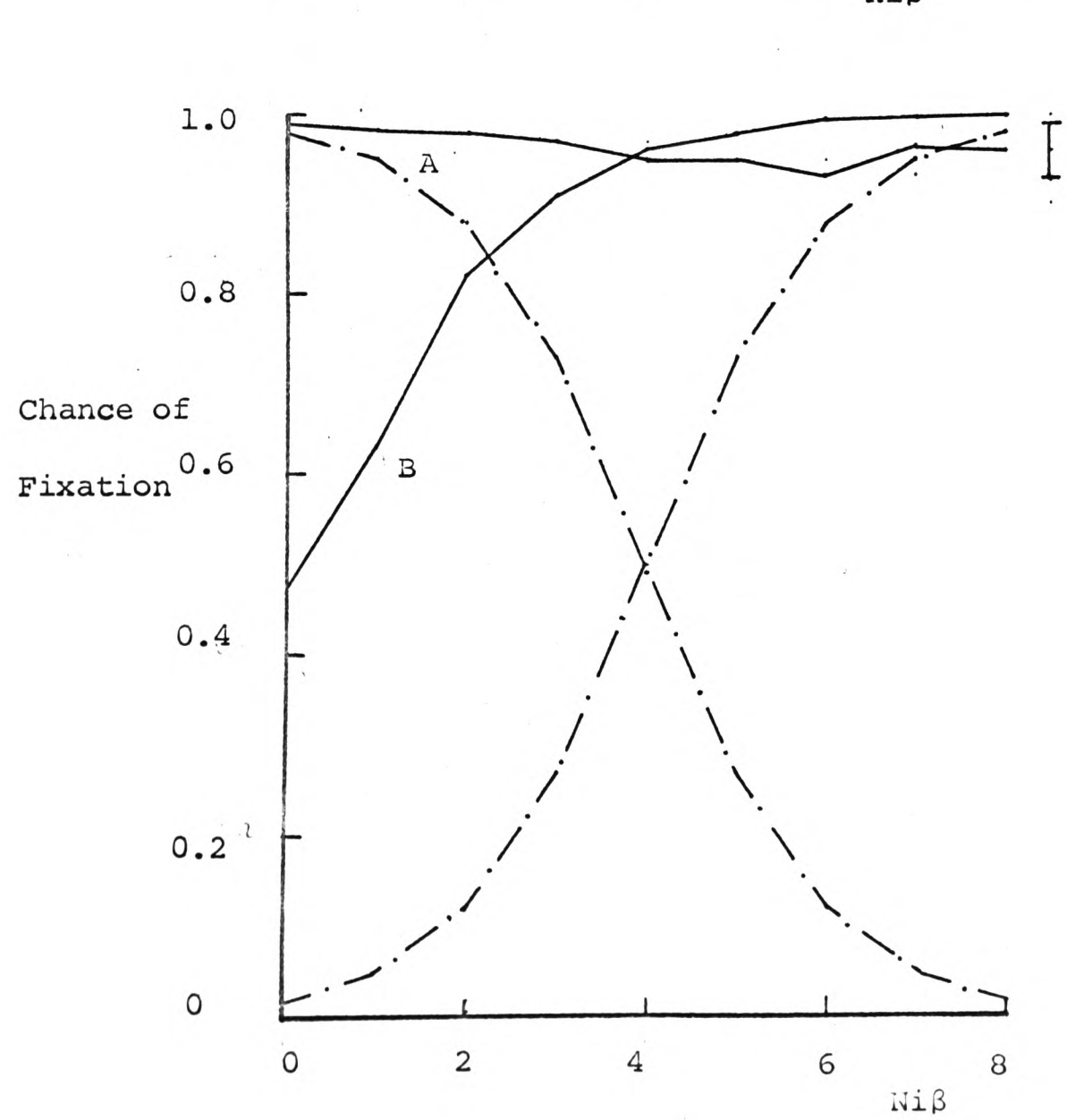


FIGURE 4.7. The influence of relative magnitude of gene effects on the importance of initial linkage disequilibrium for N=8, n=2, Niα=4, Nc=0, p=q=0.5. Typical range of length four standard errors is also shown.

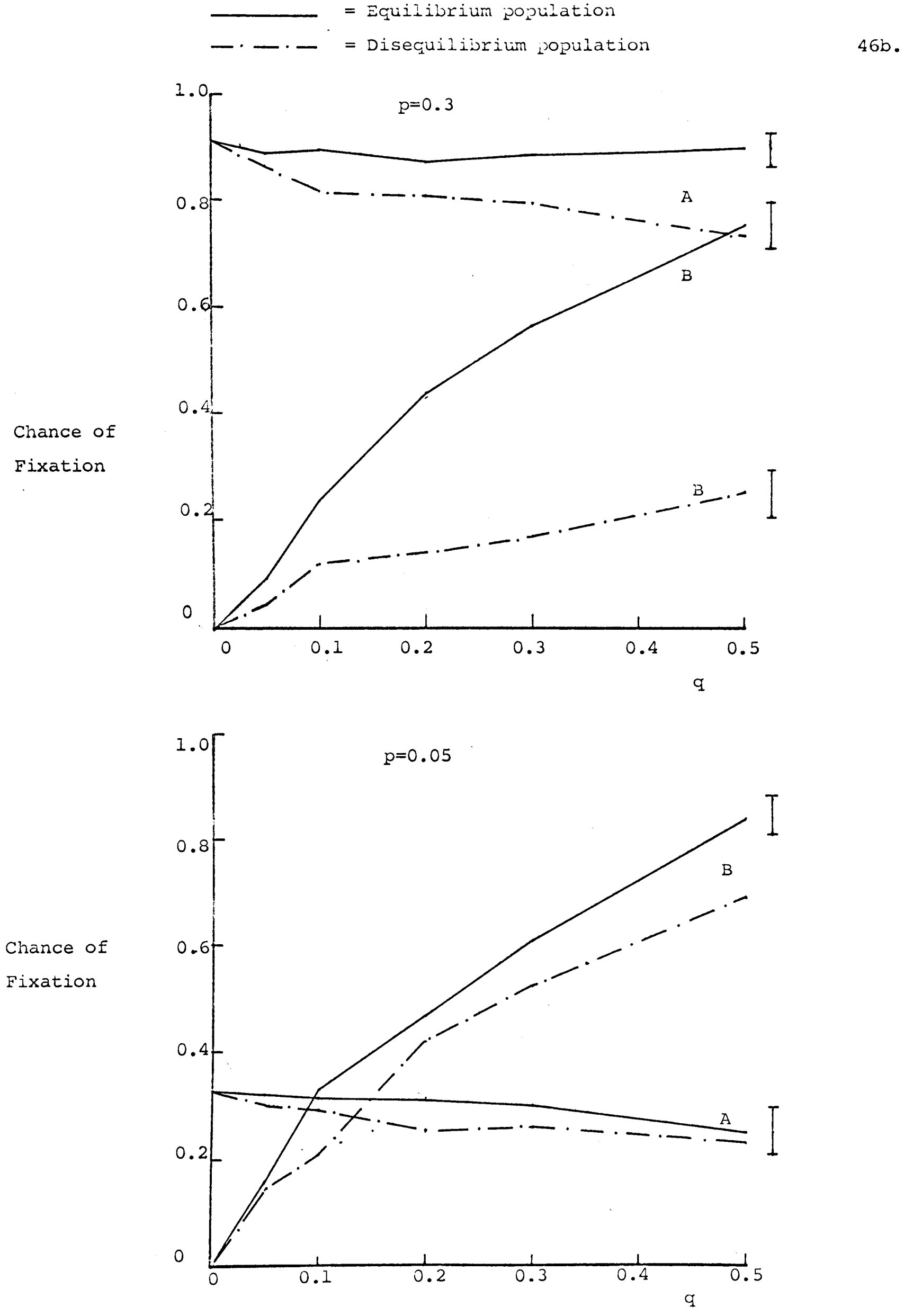


FIGURE 4.8. The effect of initial linkage disequilibrium on chance of fixation of favourable alleles for N=8, n=2, Niα=4, Niβ=2, Nc=0 and various values of p and q. Typical ranges of length four standard errors are also shown.

allele even if it is at a considerably higher frequency than the larger. The difference in M.A.R. is shown plotted against pq in figure 4.9 and again some indication of a curvilinear relationship emerges. However with unequal effects there is a difference, in terms of response at least, between the situations

i) $p = x$, $q = y$

and ii) $p = y$, $q = x$ $\quad x \neq y$

although pq is the same for both. To see if this affected the difference in M.A.R. between the disequilibrium and equilibrium populations some computer runs were done for $N = 8$, $Ni\alpha = 4$, $Ni\beta = 2$, $Nc = 0$ with $x=0.5$ and $y = 0.05$ and 0.3. No significant differences were found suggesting that the difference in M.A.R. was never-the-less still determined by the product pq.

The effect of initial linkage disequilibrium when $Nc = 0$ can be summarised as follows:

(a) the difference in M.A.R. has been found to increase with pq being near to zero for low pq and reaching a maximum of 0.5 when $p=q=0.5$ and $\alpha=\beta$

(b) the difference in M.A.R. has been found to be greatest when effects are equal and large, decreasing to zero as either or both alleles become small.

(c) With equal effects the reduction in chance of fixation has been found to be greatest for the locus with the lowest frequency favourable allele

(d) With unequal effects the reduction in chance of fixation due to disequilibrium has been found to be greatest for the

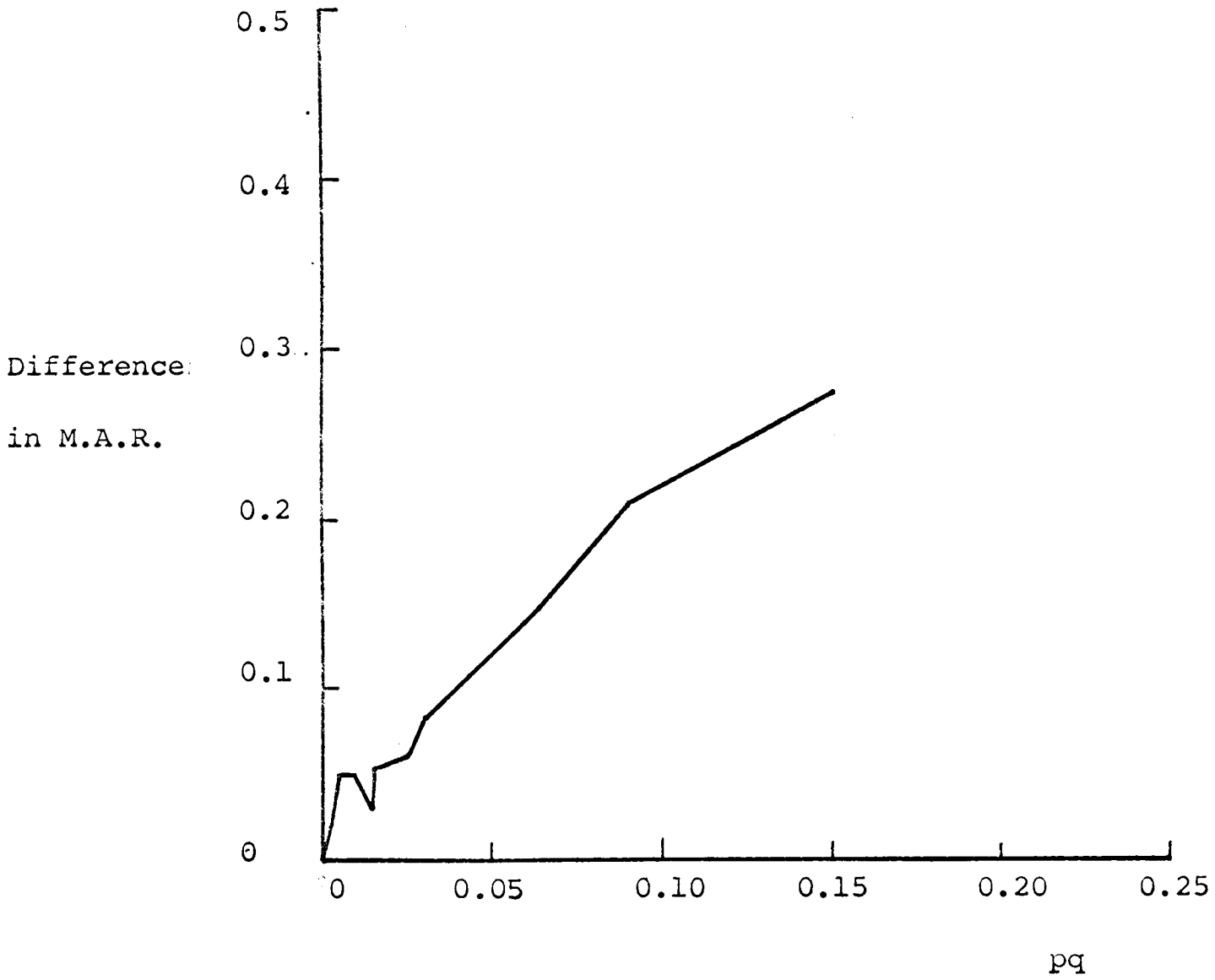


FIGURE 4.9 The relationship between the difference in mean absolute response and pq for N=8, n=2, Niα=4 Niβ=2, Nc=0

locus with the smaller effect, even if it is at a considerably higher frequency than the locus with the larger effect.

iii) Intermediate Recombination

Consider the extreme case where $p = q = 0.5$ and $\alpha = \beta$. Earlier examination of this situation showed $u(f_1)_D$ to be a function of $\frac{Nc}{2(1+2Nc)}$. In the equilibrium population $u(f_1)$ is very little affected by linkage since frequencies are high, therefore the difference in M.A.R. will be also a function of $Nc/2(1+2Nc)$. Figure 4.10 shows results for $N = 8$ and 16, and also shows how Nc varies with $Nc/2(1+2Nc)$. Therefore a small increase in Nc causes a considerable reduction in the effect of disequilibrium under these extreme conditions at least.

The conclusions reached for the no recombination case are re-examined below for intermediate recombination, to see if they still apply.

(a) Is the difference in M.A.R. still dependent on gene frequency only as a function of pq? Simulation results for $Nc=\frac{1}{4}$ suggest that this is possibly no longer true at least when p and q are small, see Figure 4.11. This may be due to chance loss of the Ab or aB gametes before recombination can occur and this will be a function of the individual gene frequencies.

(b) Is the difference in M.A.R. still maximized for equal large effects? Simulation results for a fixed α and variable β show that recombination tends to reduce the effect of the disequilibrium where it was greatest, so that although the difference in M.A.R. is still maximized in the region of $\alpha = \beta$ the curve has

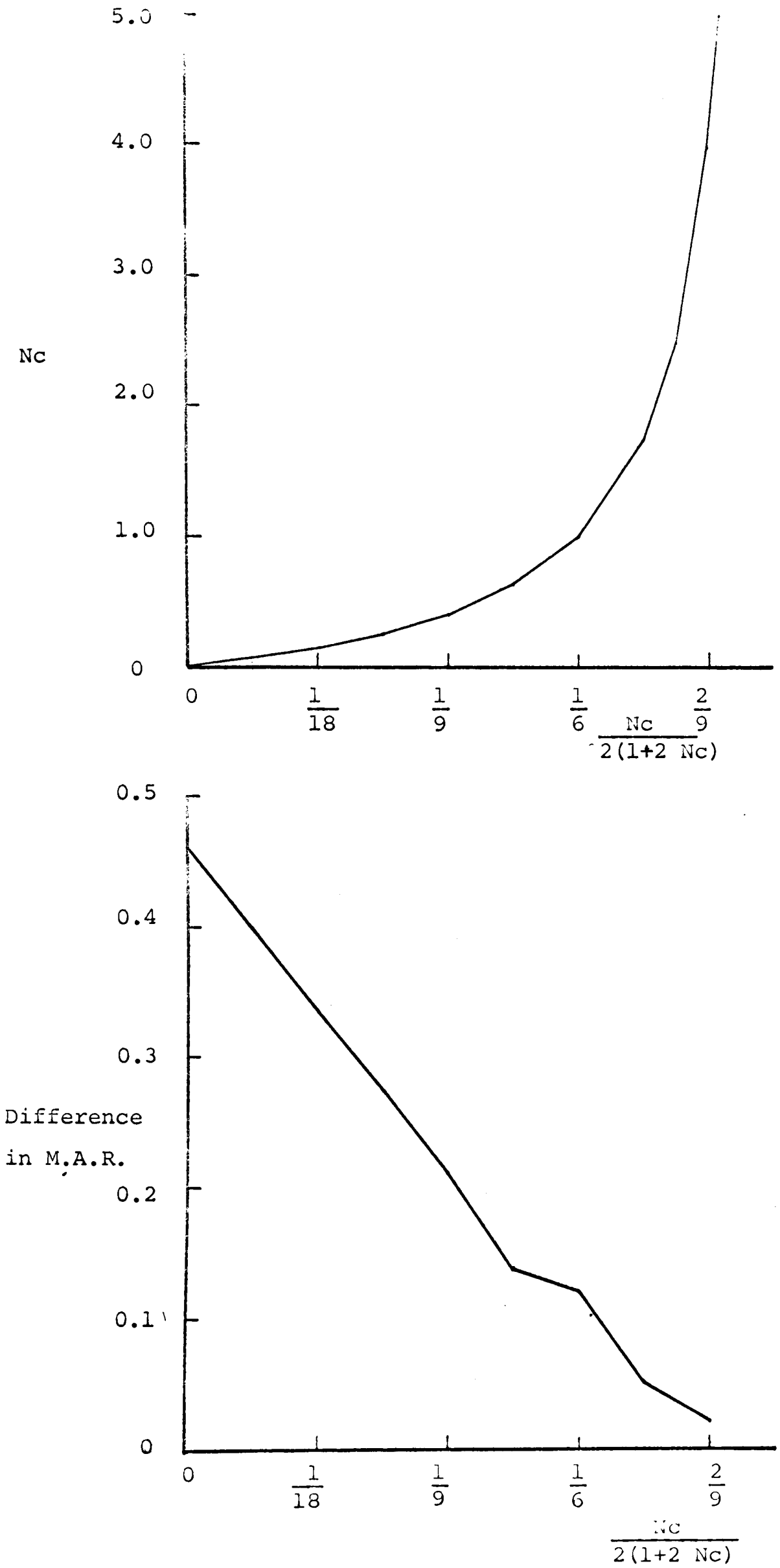


FIGURE 4.10 The relationship between the difference in mean absolute response and Nc/2(1+2Nc) for N=8, n=2, Niα=Niβ=4, p=q=0.5 with the relationship berween Nc and Nc/2(1+2Nc) also shown.

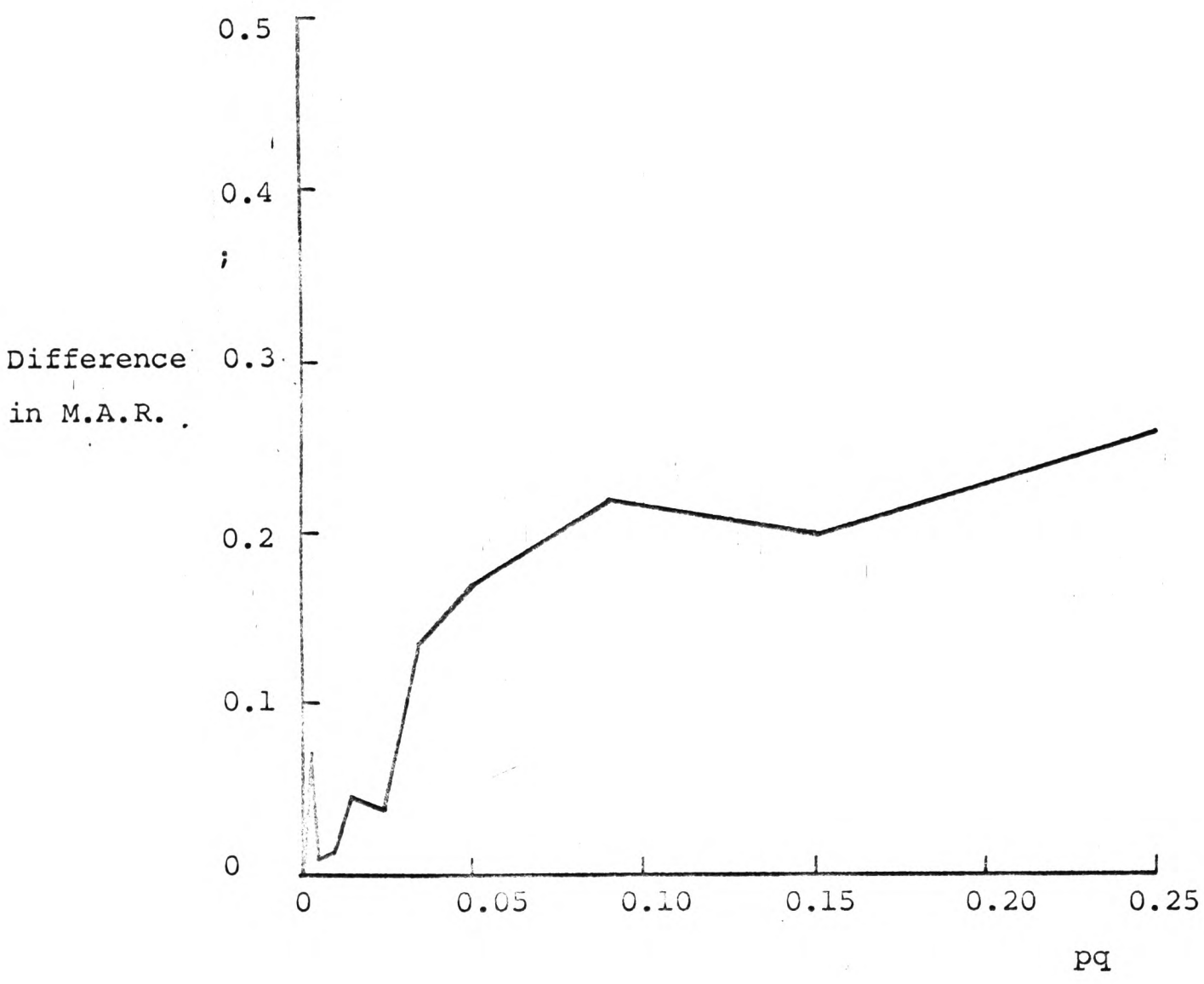


FIGURE 4.11 The relationship between the difference in mean absolute response and pq for N=8, n=2, Nc=0.25.

become much more flat-topped, see figure 4.12.

With recombination the AB gamete can be formed and so $u(p)_D$ plus $u(q)_D$ no longer reach a maximum value of one. Therefore as α and β increase $u(p)_D$ and $u(q)_D$ approach 1. Figure 4.13 shows simulation results for $p = q = 0.5$. Since $u(p)_D$, $u(q)_D$, $u(p)_E$ and $u(q)_E$ all approach one for large $N_i\alpha$ the difference in M.A.R. is maximized for an intermediate value, in this case for $N_i\alpha = N_i\beta = 6$.

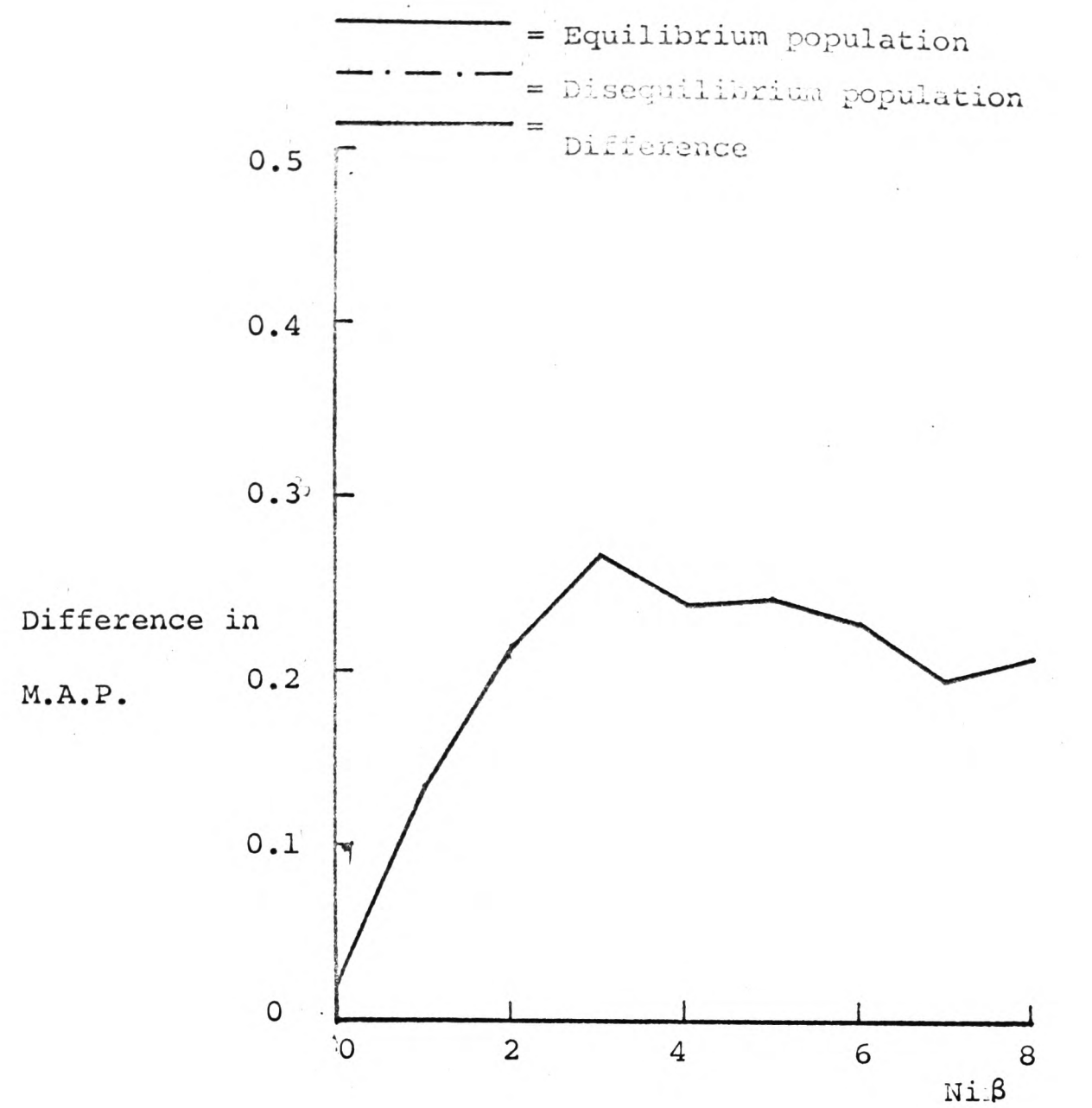
(c) With equal effects is the reduction in chance of fixation due to disequilibrium greatest for the lower frequency allele? Even with recombination equation (38) still holds although $u(f_1)_D$ is not zero so that the differences in chance of fixation are given by

$$u(p)_E - u(p)_D = u(f_1)_E - u(f_1)_D - \frac{p}{p+q}\,(u(f_2)_D - u(f_2)_E + u(f_3)_D - u(f_3)_E)$$

$$u(q)_E - u(q)_E = u(f_1)_E - u(f_1)_D - \frac{q}{p+q}\,(u(f_2)_D - u(f_2)_E + u(f_2)_D - u(f_2)_E)$$

with $u(f_2)_D > u(f_2)_E$ and $u(f_3)_D > u(f_3)_E$ although differences will be reduced from the no recombination case, the effect of the disequilibrium will still be greater for the lower frequency allele.

(d) With unequal effects will the chance of fixation be reduced more for the smaller allele? Figure 4.12 shows that this is still true although recombination tends to have more effect in reducing the effect of disequilibrium where it is most important,



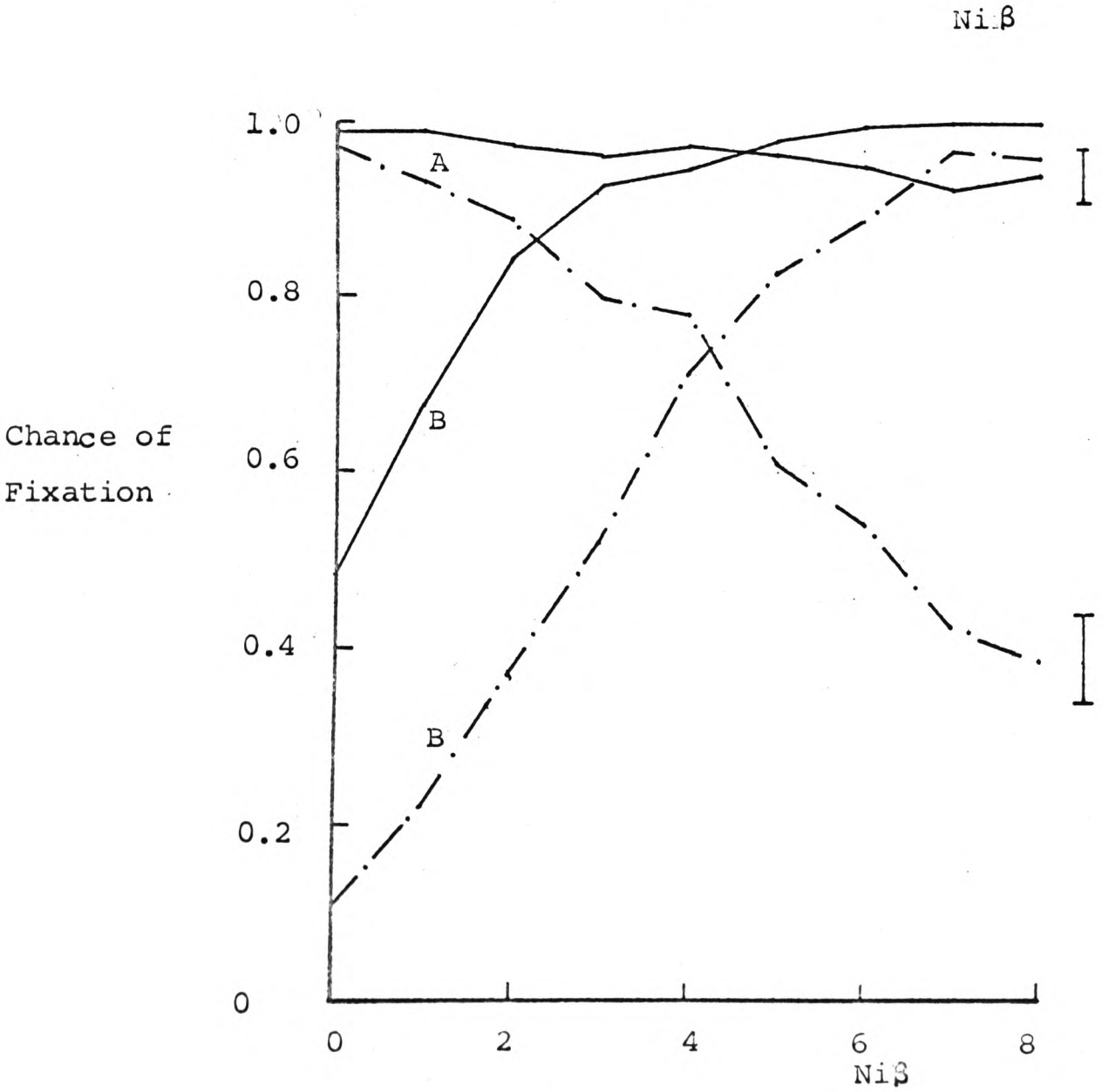


FIGURE 4.12 The influence of relative magnitude of gene effects on the importance of initial linkage disequilibrium for N = 8, n = 2, Niα = 4, Nc=0.25, p=q = 0.5. Typical ranges of length four standard errors is also shown.

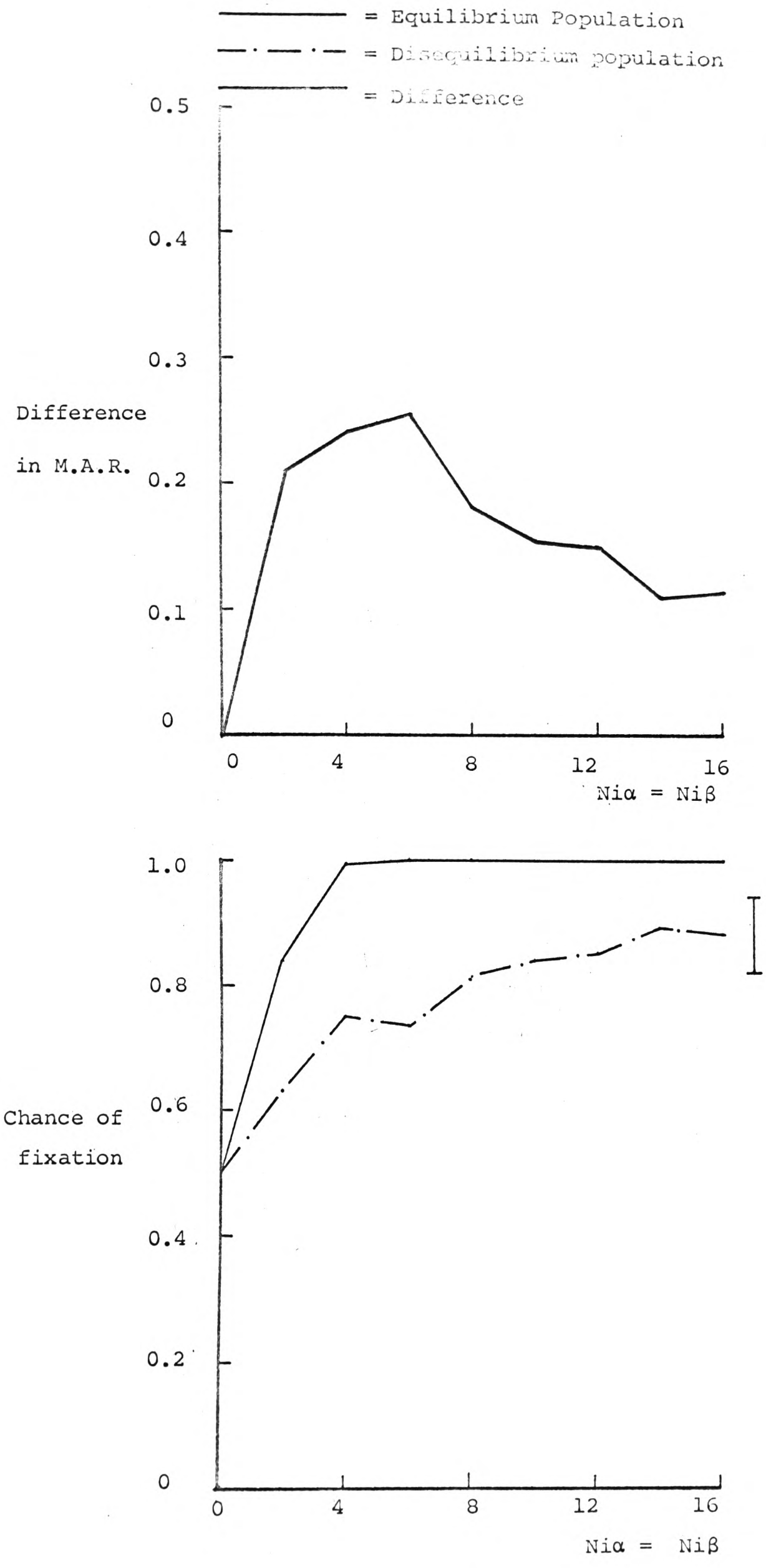


FIGURE 4.13. The influence of the absolute manitude of gene effects on the importance of initial linkage disequilibrium for N=16, n=2, Nc=0.25 p = q = 0.5. Typical range of length four standard errors is also shown.

i.e. for the smaller allele.

Finally under the two locus model some less extreme disequilibrium conditions have been briefly considered, that is when p_2 and q_1 are not zero. In particular the case where $p_1 \geq p_2$ and $q_2 \geq q_1$ has been studied then $D = + \frac{1}{4}(p_1 - p_2)(q_1 - q_2)$ is always negative, or zero. Let the mean frequencies in the cross be equal at 0.5 such that $\frac{p_1 + p_2}{2} = \frac{q_1 + q_2}{2} = p = q$

and let $p_1 = q_2$ and $q_1 = p_2$ then

$$D = + \frac{1}{4}(p_1 - p_2)^2$$

In the disequilibrium population $f_1 = f_4 = p_1 p_2$. The case where $p_1 = 1$, $p_2 = 0$ has been considered above then if there is no recombination and equal effects $u(p)_D = u(q)_D = 0.5$. For p_1 slightly less than 1 the AB and ab gametes will be present at low frequency and the chance of fixation of AB can be given approximately by ignoring the presence of the ab gam ete entirely, then

$$u(f_1)_D = \frac{1-e^{-2Ni\alpha p_1 p_2}}{1-e^{-2Ni\alpha}} \qquad \ldots(40)$$

As p_1 decreases towards 0.5 the effect of the ab gamete may become appreciable and equation (4) may underestimate $u(f_1)_D$. $p_1 = p_2 = 0.5$ gives the equilibrium case. Figure 4.14 gives simulation results for $Ni\alpha = 4$ plotted against D together with the theoretical curve. This shows that the effect of initial linkage disequilibrium is rapidly reduced as D moves away from the extreme negative value of -0.25, for example a 20% reduction in the magnitude of the disequilibrium causes approximately a 40% reduction in its effect on the

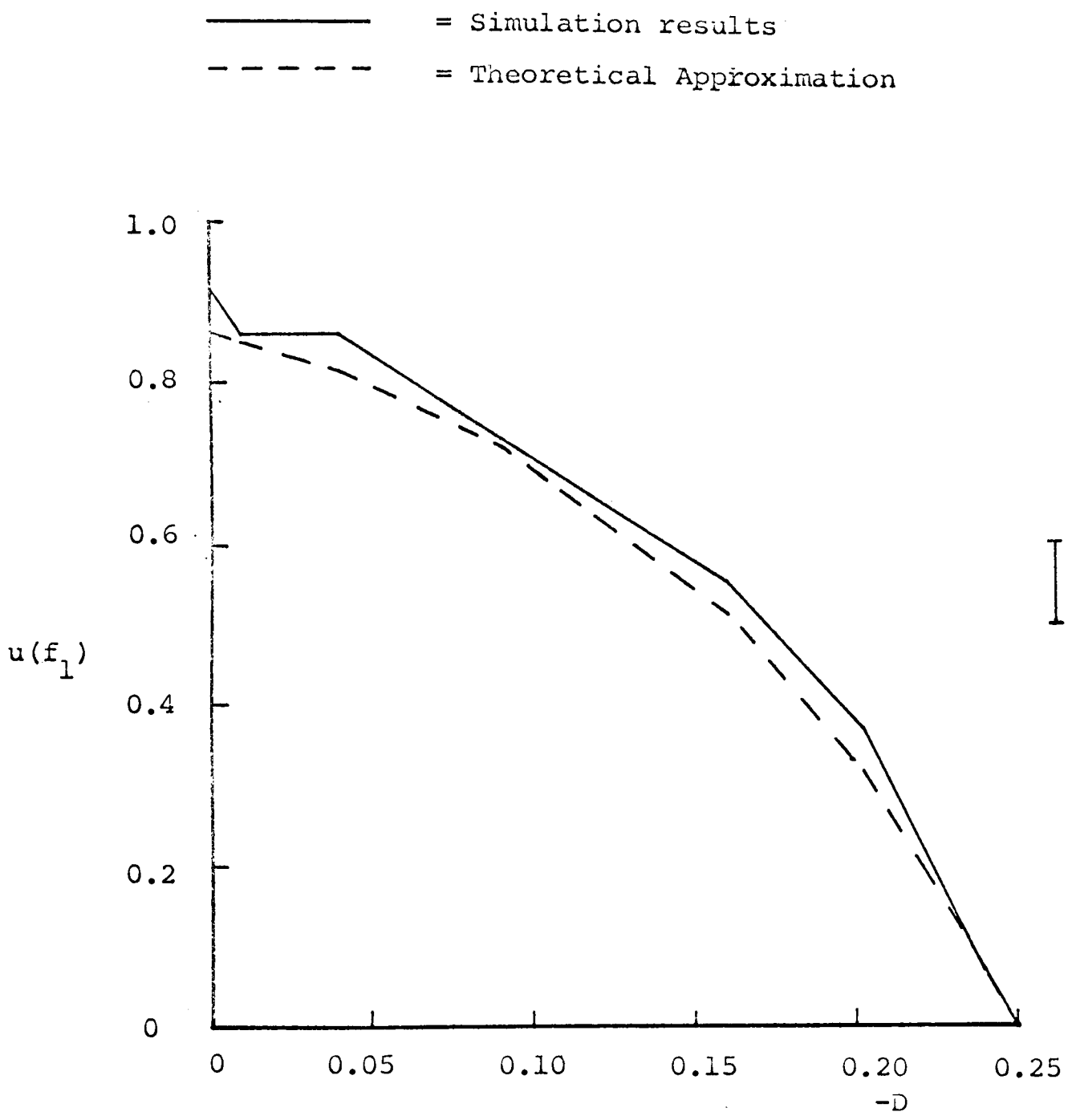


FIGURE 4.14. The effect of reducing the initial linkage disequilibrium for the same initial mean gene frequency for N=8, n=2 $Ni\alpha=Ni\beta=4$. Nc=0, $p = \frac{p_1 + p_2}{2} = q = \frac{q_1 + q_2}{2} = 0.5$ with $p_1=q_2$ and $q_1=p_2$ such that $D_o = -(p_1 - 0.5)^2$. The theoretical approximation $u(f_1) - \frac{1-e^{-2Ni\alpha pq}}{1-e^{-2Ni\alpha}}$ is shown. Typical range of length four standard errors is also shown.

response. Even for values of p_1 approaching 0.5 the theoretical approximation gives a reasonably good prediction of chance of fixation.

These investigations have shown that even for the extreme case initially considered, where D = -pq, the effect of the disequilibrium may not be very great, for example if pq is low or Nc is high. Therefore there are many situations where initial linkage disequilibrium might be ignored since its effect on the response is small. To illustrate this an arbitrary level has been chosen of a difference in M.A.R. of 0.1 and any differences below this have been assumed to be trivial. In this way the results from simulation can be summarized in table form giving the approximate values of pq for which initial linkage disequilibrium does not have an appreciable effect, see Table 1.

Table 1.

		Recombination Fraction	
Niα	Niβ	Nc = 0	0 < Nc < $\frac{1}{4}$
Large	Small	pq < 0.05	pq < 0.09
Large	Large	pq < 0.02	pq < 0.05
Small	Small	pq < 0.04	pq < 0.12

Values of pq which produce disequilibrium having negligible effect on response.

Thus if the response to selection is to be maximized for a population in linkage disequilibrium it may in some circumstances be

worth trying to restore equilibrium before selection is commenced. However in order to restore a disequilibrium population to equilibrium some degree of recombination is required. (This is not strictly true for the case where the AB gamete is present although at a reduced frequency, then some special form of selection might restore equilibrium but this will not be considered here.) So for those cases where Nc = 0 there is no possibility of attaining equilibrium. The conditions under which it might be worth attempting to restore equilibrium can be summarized as

$$0 < Nc \leqslant \tfrac{1}{4}\ , \quad pq \geqslant 0.03\ , \quad \frac{\alpha}{2} \leqslant \beta \leqslant 2\alpha \text{ for } \alpha \text{ given by}$$

$2 \leqslant Ni\alpha \leqslant 16$.

The next question then to be considered is the way in which linkage disequilibrium might be broken down.

(c) <u>The Break-down of Linkage Disequilibrium</u>

Consider the following situations:

i) Infinite population, no selection

ii) Infinite population with selection for additive loci

iii) Finite population, no selection

iv) Finite population with selection for additive loci.

i) Infinite population, so selection.

In general

$$D_{t+1} = (1 - c)\, D_t \qquad \ldots(41)$$

If the first generation is a random cross

$$D_t = (1 - c)^t D_o$$

If the first generation is an F_1 cross

$$D_t = (1 - 2c)\ (1 - c)^{t-1} D_o$$
$$\approx (1 - c)^{t+1}\ D_o \text{ if c is small.}$$

ii) Infinite population with selection.

In general

$$D_{t+1} = D_t(1 - c)\ (1 - \frac{i}{2}\ (\alpha\ + \beta - 2\mu)) - \Delta p \Delta q \quad \ldots(42)$$

where Δp and Δq are the changes in gene frequency due to selection. Figure 4.15 shows how selection increases the rate of breakdown of the disequilibrium particularly if the effects α and β are unequal. In that case even for c = 0 the disequilibrium is rapidly broken down, however this is not due to any change in the ratio of frequencies of repulsion and coupling phase gametes but simply to the change in the values of f_2 and f_3. Hill and Robertson (1968) suggested using the square of the correlation of gene frequencies as a measure of the disequilibrium such that

$$r^2 = \frac{D^2}{pq(1-p)(1-q)} \qquad 0 < p\ q < 1 \qquad \ldots(43)$$

If $f_1 = f_4 = 0$, $q = 1 - p$ and $r^2 = 1$ for all p. However with f_1 and f_4 not zero the value which r^2 can take is not independent of the marginal gene frequencies, this fact was pointed out by Sved (1971).

iii) Finite population, no selection.

In general

$$D_{t+1} = (1 - c)\ (1 - \frac{1}{2N})\ D_t \qquad \text{Hill and Robertson} \ldots(44)$$
$$D_{t+1} \approx (1 - c - \frac{1}{2N})\ D_t$$

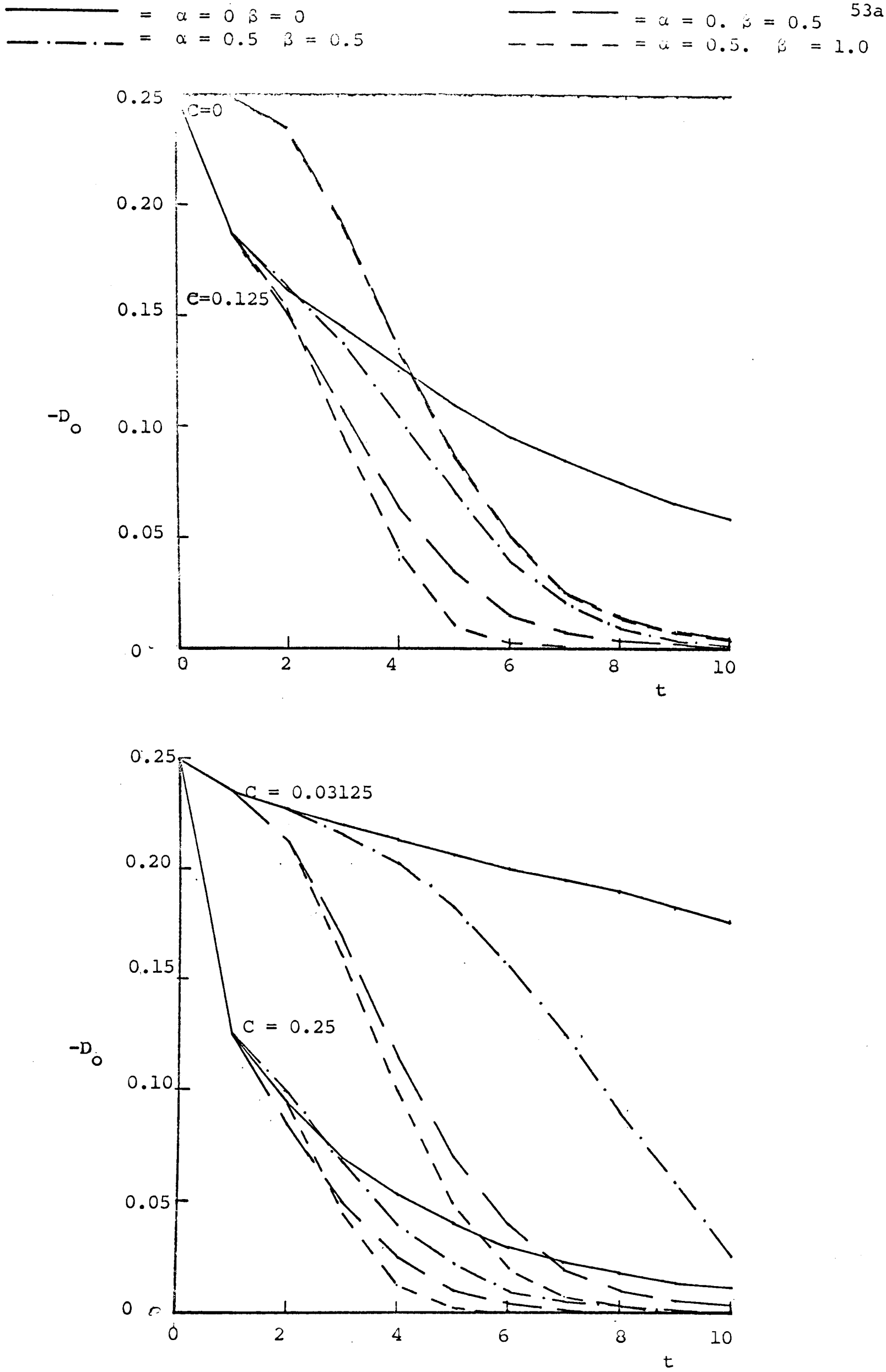


FIGURE 4.15 The rate of break-down of initial linkage disequilibrium with an F_1 cross for $N=\infty$, $n=2$, $p=q=0.5$, $D_o = -0.25$, with various values of α, β and c.

If c is small such that $c/_{2N}$ can be ignored.

Therefore a reduction in population size tends to increase the rate of breakdown of the disequilibrium. This is due to random changes in gene frequency leading eventually to fixation of one or other allele at each locus, as this occurs, so D approaches zero.

iv) Finite population, with selection.

No general term is known for this case but from ii) above selection would be expected to increase the rate of breakdown of the disequilibrium, particularly if effects were unequal. However the effect of drift in a small population must be considered, this tends to reduce the average rate of response to selection by reducing the rate of gene frequency changes. Therefore with selection operating the introduction of drift might be expected to reduce the rate of breakdown of the disequilibrium when compared with the infinite population. Simulation results for no selection, selection with equal effects and selection with unequal effects are shown in Figure 4.16, for $D_o = -0.25$. Differences in the rate of breakdown are only found for $Nc < \frac{1}{4}$ and then only with selection for loci of unequal effect, in which case selection increases the rate of breakdown. Therefore it appears that the effect of drift obliterates all but the largest of the differences found for the infinite population case.

If linkage disequilibrium is to be reduced towards zero prior to selecting from a cross this can only be achieved without altering gene frequencies if a period of relaxation in an infinite population is permitted. To see if this might be useful consider

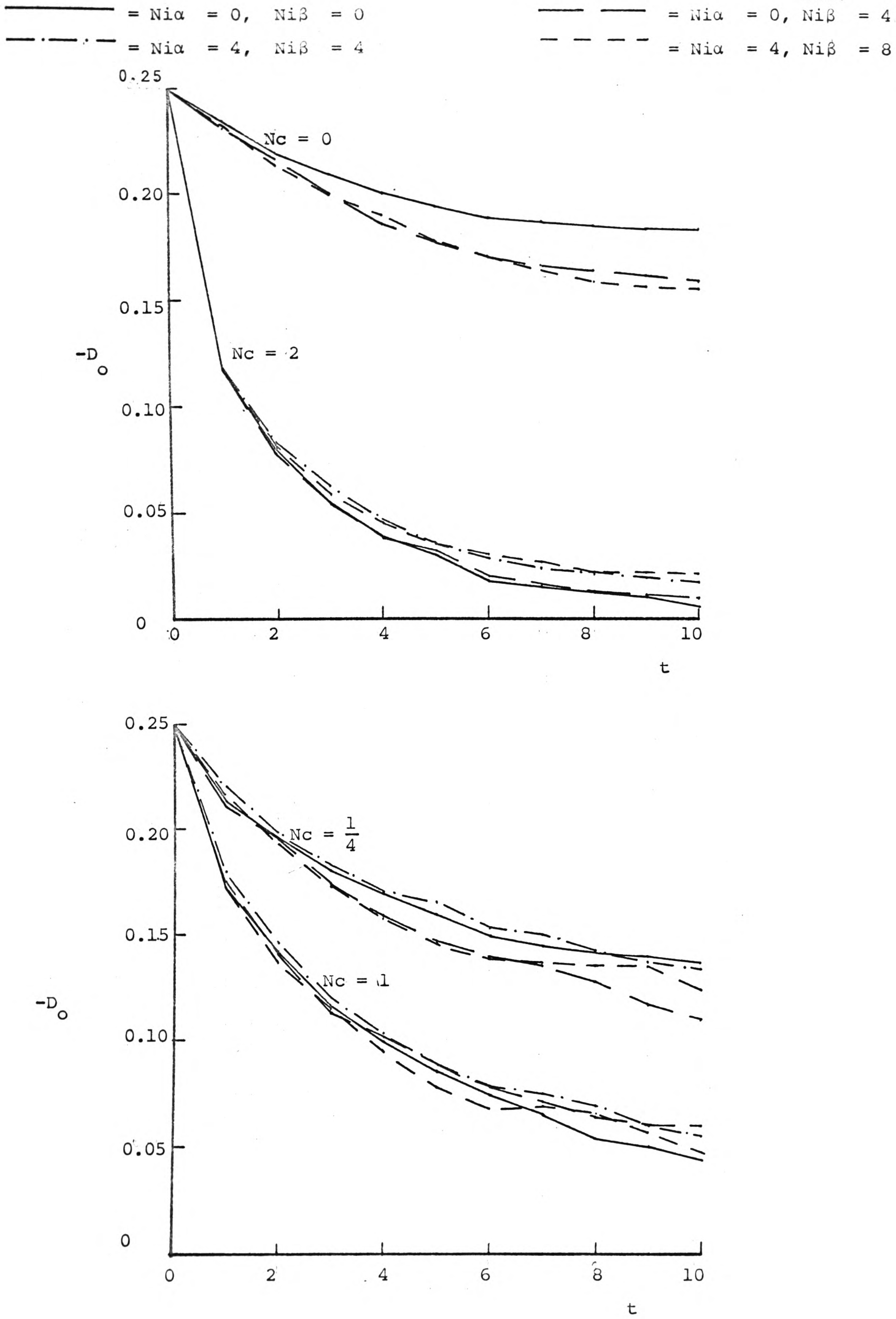


FIGURE 4.16. The rate of breakdown of initial linkage disequilibrium with an F_1 cross, for N = 8, n = 2, p = q = 0.5. D_o = -0.25 with various values of α, β and c.

how long it would take to reduce the disequilibrium by 20% this is given by

$$t \approx \frac{\log (0.8)}{\log (1-c)} - 1 \qquad \text{for the } F_1 \text{ cross case}$$

e.g. for $c = 0.03125$ $\quad t \approx 6$ generations

so that a relatively short period of relaxation in an infinite population may make an appreciable difference in the disequilibrium. However only if c is very small will such a reduction be useful and in that case the time required may be considerably longer.

The next point to consider is just how feasible such a period of relaxation is. An effectively infinite population may be kept unselected in some cases, for example experimental animals such as Drosophila or many plant species, but in others, such as large domestic animals, this is not a practical possibility. In view of this the effect of a period of relaxation in a finite population has been examined. As has already been pointed out this involves random changes in gene frequencies and this increases the rate of breakdown of the disequilibrium but because of the changes in gene frequencies this situation is not directly comparable with the infinite population case. Restriction of the population size during relaxation also allows for chance loss of favourable gametes and in this respect it may be detrimental so that the population size should be kept as large as possible, say M. Simulation has been carried out to observe the effect of T generations of relaxation in a population of size M, on the ultimate response given by the M.P.R. and on the chance of fixation of the AB gamete. Figure 4.17 shows results for a) equal

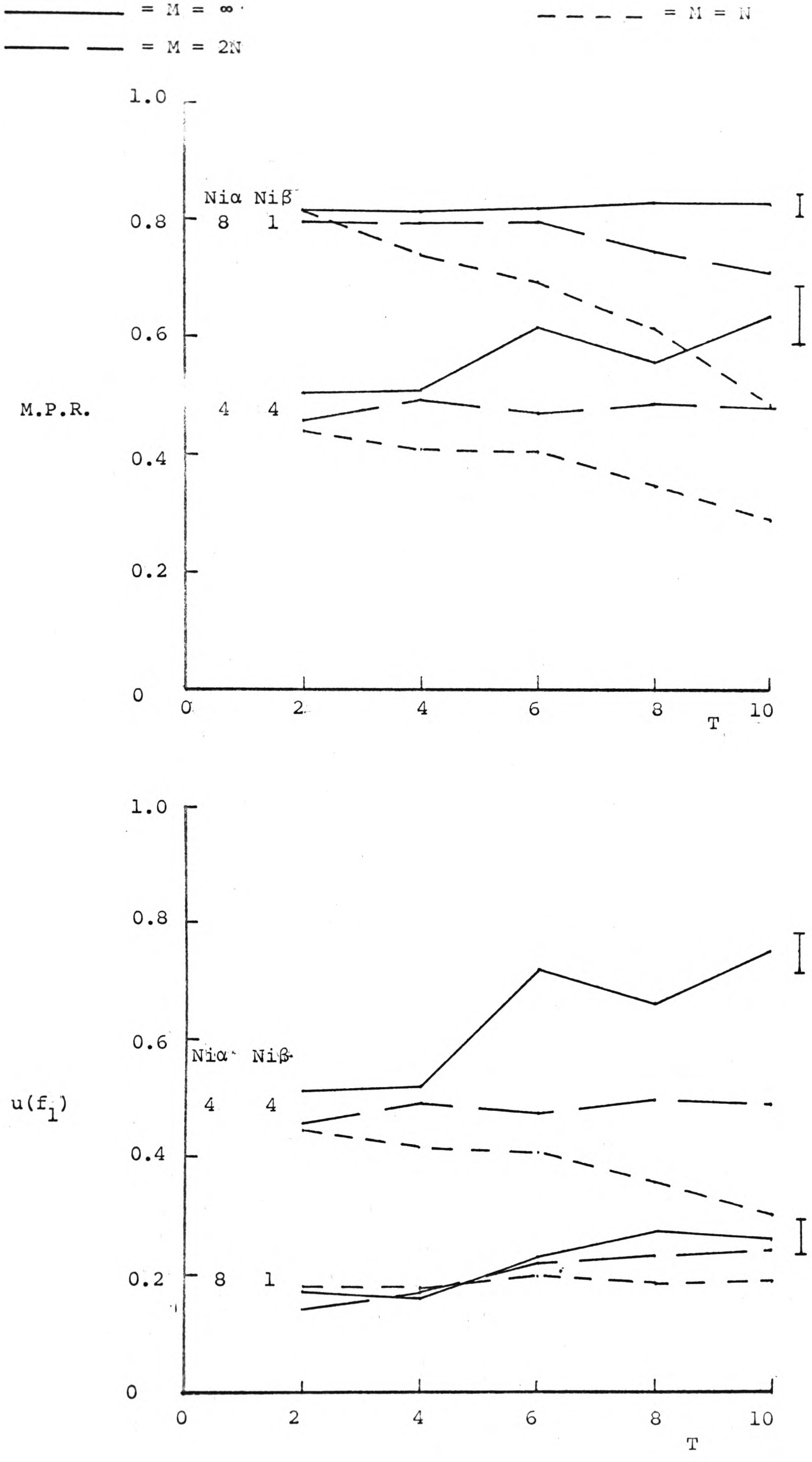


FIGURE 4.17 The effect of T generations of relaxation before selection for N = 8, n = 2, Nc = 0.25, p = q = 0.5, D_o = -0.5, with various values of Niα , Niβ and M. Typical ranges of length four standard errors are also shown.

effects and b) unequal effects with $D_o = -0.25$ and Nc = ¼ with M taking values of N, 2N and ∞. With equal effects M = ∞ increases the M.P.R. and chance of fixation of AB, M = 2N makes little difference to either and M = N decreases both. However with unequal effects the M.P.R. is barely increased for M = ∞ and is decreased for M = 2N and M = N, while the chance of fixation of AB is increased for all cases. Therefore a period of relaxation in a finite population is unlikely to be of any appreciable benefit in terms of mean response but it may be useful in incorporating a smaller locus into a population already fixed for one locus of large effect, or several loci of small effect.

Some attempt has been made to consider the effect on response of a period of relaxation under a multilocus model but discussion of these results will be deferred until after the measurement and effect of disequilibrium under such a model has been considered.

(d) <u>The Measurement of linkage disequilibrium for many loci.</u>

The linkage disequilibrium between any pair of loci is given by

$$D_{ij} = f_{ij} - p_i p_j \qquad \ldots(45)$$

where f_{ij} = the frequency of the gamete carrying favourable alleles at the ith and jth loci.

The disequilibrium between any number of loci can be formulated in the same way but for non-epistatic loci it is only the disequilibrium between pairs which enters into equations predicting gene

frequency changes and additive genetic variances in finite populations. For this reason only the disequilibrium between pairs will be considered here.

For any population where gamete and gene frequencies are known the values of the $n(n-1)/2$ D_{ij} values may be calculated (where n is the number of loci), the problem then becomes one of considering all these values to arrive at some general measure of the linkage disequilibrium for the population. For example consider an extreme case of an F_1 cross between two selected lines which have become fixed. All loci segregating in the cross will be at a frequency of 0.5 and the disequilibrium between each pair must be $\pm$ 0.25. E.g. for a six locus model let the genotype be

Chromosome	A	B	C	D	E	F	
Frequency in Pop^n 1	1	0	1	0	1	0	
Frequency in Pop^n 2	0	1	0	1	0	1	genotype (a)

then $D_{AB} = D_{BC} = D_{CD} = D_{DE} = D_{EF} = D_{AD} = D_{BE} = D_{CF} = D_{AF} = -0.25.$

$$D_{AC} = D_{BD} = D_{CE} = D_{DF} = D_{AE} = D_{BF} = +0.25$$

Let $\overline{D} = \sum_{i} \sum_{j \neq i} D_{ij} / (n(n-1)/2)$

in this case $\overline{D} = -1/20$

This will be termed the unweighted mean disequilibrium and it is the same for all F_1 populations in which each chromosome of the pair carries three favourable and three unfavourable alleles. Another genotype with the same $\overline{D}$ is given by:

Chromosome	A	B	C	D	E	F	
Frequency in Pop^n 1	1	1	1	0	0	0	genotype (b)
Frequency in Pop^n 2	0	0	0	1	1	1	

So although genotypes (a) and (b) have the same $\bar{D}$ they are quite different in arrangement. Clearly if there is no recombination between any of the pairs of loci then order is of no importance, this is also true for the F_1 case if there is free recombination. However with intermediate recombination the differences in order between genotypes (a) and (b) would be expected to have a marked effect on response to selection and ultimate chance of fixation of the loci. Therefore some measure is required which will take order into account, Fraser, Miller and Burnell (1965) devised such a system defining 'Potency' as the number of 1s on the chromosome and "Recombination index" as the number of changes from 0 to 1 (or 1 to 0) going along the chromosome. In this terminology genotype (a) has potency 3, recombination index 5, while genotype (b) has potency also 3 but recombination index 1. This system will be considered further in the light of simulation results. First two alternative systems will be outlined and discussed, these are possible ways of weighting the mean disequilibrium to take order into account.

I) Consider the effect of one generation of recombination on the linkage disequilibrium. If there was an F_1 cross, as is assumed here, then

$$D_{ij}^{1} = (1 - 2c_{ij})D_{ij}^{0}$$

Let D_{wl} = the expectation of $\bar{D}$ after 1 generation of recombination

$$= \sum_{i \neq j} (1 - 2c_{ij}) D_{ij}/(n(n-1)/2) \qquad \ldots(46)$$

e.g. compare (a) 1 1 0 with (b) 1 0 1
0 0 1 0 1 0

for (a) $D_{wl} = -\frac{1}{4} \times \frac{1}{3} - (-c + c^2)/3$

$= (-1 + 4c - 4c^2)/12$

where c = the recombination fraction between each adjacent pair of loci,

for (b) $D_{wl} = -\frac{1}{4} \times \frac{1}{3} - (-c^2)/3$

$= (-1 + 4c^2)/12$

If $c = 0$, D_{wl} for (a) and (b) $= -1/12$

If $c = 0.5$, D_{wl} for (a) and (b) $= 0$

But for $0 < c < 0.5$ $|D_{wl}|_{(a)} < |D_{wl}|_{(b)}$

II) An alternative approach to this problem is to weight the disequilibria between pairs directly by some function of the distance between them. Clearly linkage disequilibrium between very tightly linked loci is of more importance than that between more distant loci, consequently the reciprocal of the map distance between members of each pair might give an appropriate weighting factor.

Let r = the map distance between each adjacent pair of loci, and

r_{ij} = the map distance between the ith and jth loci, then let

$$D_{w2} = \sum_{i \neq j} D_{ij} \times \frac{1}{r_{ij}} / (n(n-1)/2) \qquad \ldots(47)$$

so for (a) 1 1 0
0 0 1

$$D_{w2} = [\frac{1}{4} \times \frac{1}{r} + -\frac{1}{4} \times \frac{1}{r} + -\frac{1}{4} \times \frac{1}{2r}] \times \frac{1}{3}$$

$$= -\frac{1}{2r} \times \frac{1}{12}$$

for (b) 1 0 1

0 1 0

$$D_{w2} = [-\frac{1}{4} \times \frac{1}{r} + -\frac{1}{4} \times \frac{1}{r} + \frac{1}{4} \times \frac{1}{2r}] \times \frac{1}{3}$$

$$= -\frac{3}{2r} \times \frac{1}{12}$$

For $r = \infty$ D_{w2} for (a) and (b) = 0

$r = 0$ D_{w2} for (a) and (b) = ∞

for $0 < r < \infty$ $|D_{w2}|_{(a)} < |D_{w2}|_{(b)}$

Consider the relationship between D_{w1} and D_{w2}

$$D_{w1} = \sum_{i \neq j} D_{ij} (1 - 2c_{ij}) / (n(n-1)/2)$$

$$= \sum_{i \neq j} D_{ij} \frac{1}{e^{2r_{ij}}} / (n(n-1)/2)$$

Therefore $$\frac{D_{w1}}{D_{w2}} = \frac{\sum_{i \neq j} D_{ij}/e^{2r_{ij}}}{\sum_{i \neq j} D_{ij}/r_{ij}} \quad \ldots(48)$$

Computer simulation studies have been made to compare the relationship between linkage disequilibrium and chance of fixation for D_{w1} and D_{w2}.

For a five locus model all possible chromosome arrangements were compared with the appropriate values of D_{w1} and D_{w2}. Results are shown in Table 2, for clarity both equations and values of D_{w1} and D_{w2} have been multiplied throughout by a factor of 40 to remove the denomimator term, actual values of D_{w2} given have been further multiplied by 1.2r to simplify the calculations. Figure 4.18 shows scattergrams of u(q) against D_{w1} and D_{w2}, this indicates that there may be linear relationships within $\bar{D}$ groups although

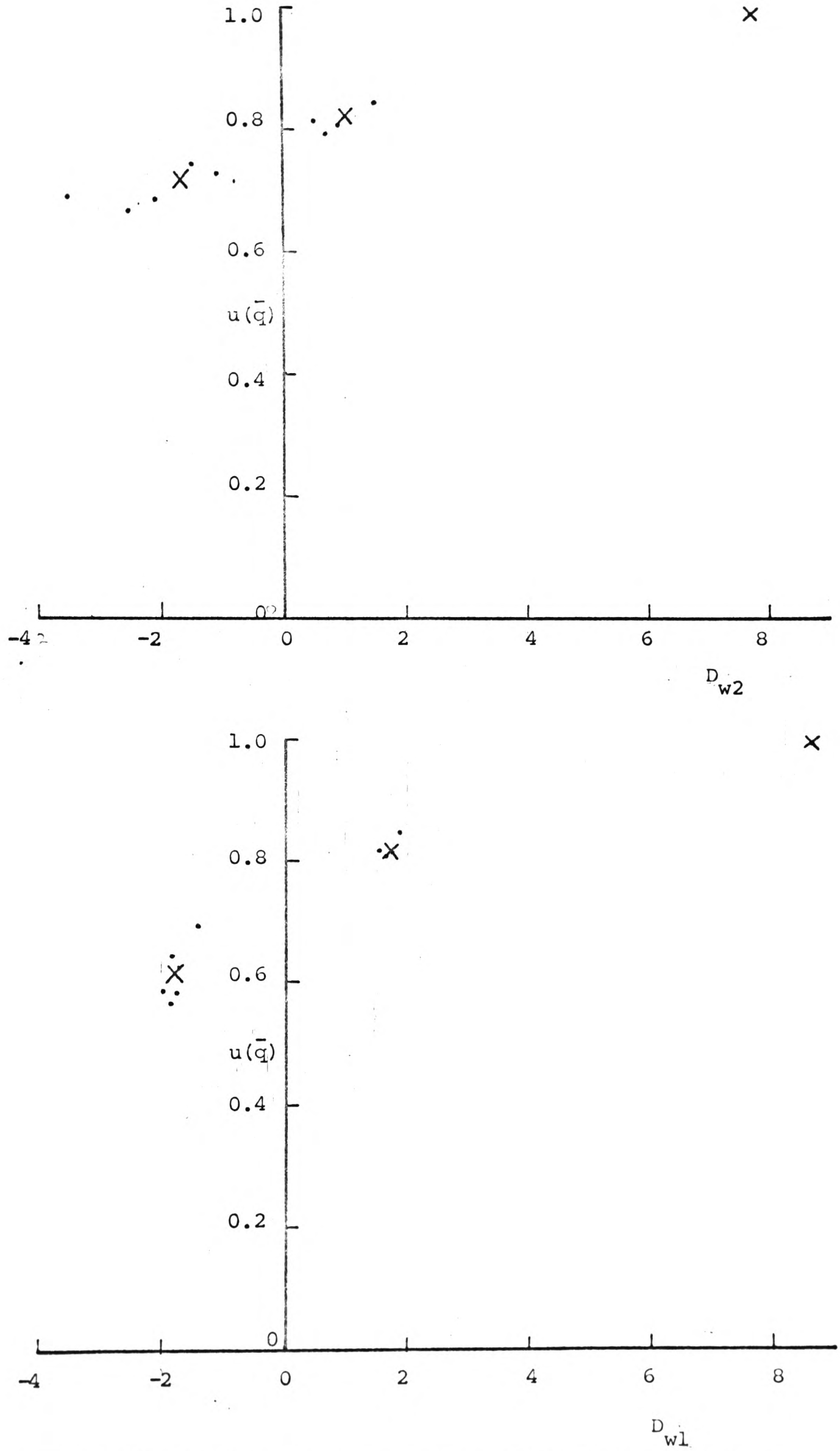


FIGURE 4.18 Scattergrams showing the relationships between chance of fixation and the parameters D_{w1} and D_{w2} for N = 10 n = 5, Niα = 2.5, Nc = 0.3125, q = 0.5. X marks the within $\bar{D}$ group mean.

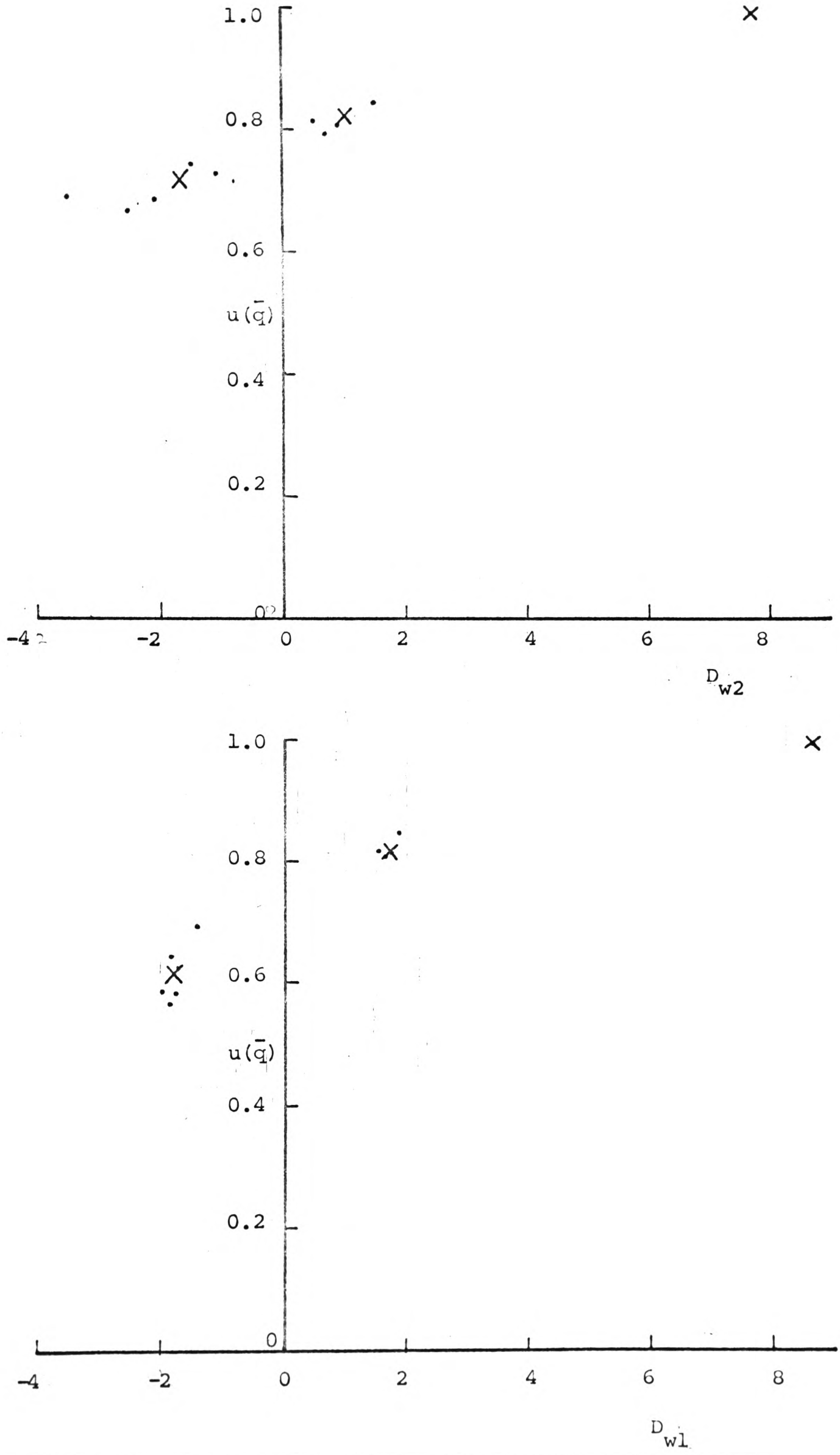


FIGURE 4.18 Scattergrams showing the relationships between chance of fixation and the parameters D_{w1} and D_{w2} for N = 10 n = 5, Niα = 2.5, Nc = 0.3125, q = 0.5. X marks the within $\bar{D}$ group mean.

TABLE 2

				c = 0.03125, r = 0.03225					
Genotype	$\overline{D}$	Equation for D_{w1}	Equation for D_{w2}	D_{w1}	$\overline{D}_{w1}$	D_{w2}	$\overline{D}_{w2}$	u(q) ±s.e.	$\overline{u(q)}$ ±s.e.
1 1 1 1 1 0 0 0 0 0	+10	$10 - 40c + 60c^2 - 48c^3 + 16c^4$	$\frac{4}{r} + \frac{3}{2r} + \frac{2}{3r} + \frac{1}{4r}$	8.69	8.69	7.70	7.70	0.994 ±0.0034	0.994 ±0.0034
1 1 1 1 1 0 0 0 0 0	+ 2	$2 - 20c^2 + 32c^3 - 16c^4$	$\frac{2}{r} + \frac{1}{2r} + \frac{0}{3r} - \frac{1}{4r}$	1.98		1.50		0.846 ±0.0104	
1 1 1 0 1 0 0 0 1 0	+ 2	$2 - 12c + 28c^2 - 32c^3 + 16c^4$	$\frac{0}{r} + \frac{1}{2r} + \frac{0}{3r} + \frac{1}{4r}$	1.65		0.90		0.809 ±0.0092	
1 1 0 1 1 0 0 1 0 0	+ 2	$2 - 16c + 44c^2 - 48c^3 + 16c^4$	$\frac{0}{r} - \frac{1}{2r} + \frac{2}{3r} + \frac{1}{4r}$	1.54	1.72	0.50	1.30	0.815 ±0.0103	0.823 ±0.0058
1 1 1 0 0 0 0 0 1 1	- 2	$-2 + 20c - 52c^2 + 48c^3 - 16c^4$	$\frac{2}{r} - \frac{1}{2r} - \frac{2}{3r} - \frac{1}{4r}$	-1.42		0.70		0.794 ±0.0192	
1 0 1 1 0 0 1 0 0 1	-2	$-2 + 4c - 4c^2 - 16c^3 - 16c^4$	$-\frac{2}{r} - \frac{1}{2r} + \frac{2}{3r} - \frac{1}{4r}$	-1.88		-2.50		0.668 ±0.0123	
1 1 0 1 0 0 0 1 0 1	-2	$-2 + 8c - 20c^2 + 32c^3 - 16c^4$	$-\frac{2}{r} + \frac{1}{2r} + \frac{0}{3r} - \frac{1}{4r}$	-1.77		-2.10		0.685 ±0.0143	
0 1 1 1 0 1 0 0 0 1	-2	$-2 + 8c - 4c^2 - 16c^3 + 16c^4$	$\frac{0}{r} - \frac{1}{2r} - \frac{2}{3r} + \frac{1}{4r}$	-1.75		-1.10		0.727 ±0.0141	
1 1 0 0 1 0 0 1 1 0	-2	$-2 + 4c + 12c^2 - 32c^3 + 16c^4$	$\frac{0}{r} - \frac{3}{2r} + \frac{0}{3r} + \frac{1}{4r}$	-1.86		-1.50		0.742 ±0.0156	
1 0 1 0 1 0 1 0 1 0	-2	$-2 + 12c^2 - 16c^3 - 16c^4$	$-\frac{4}{r} + \frac{3}{2r} - \frac{2}{3r} + \frac{1}{4r}$	-1.99	-1.78	-3.50	-1.67	0.688 ±0.0133	0.717 ±0.0056

cont•••

TABLE 2 continued

$u(q)$ = the mean chance of fixation in a simulated population of size $N = 10$ with $Ni\alpha = 2.5$, $q = 0.5$ for all loci and $c = 0.03125$ between each adjacent pair of loci.

$\overline{u(q)}$, $\overline{D_{w1}}$ and $\overline{D_{w2}}$ are mean values calculated within groups having the same $\overline{D}$.

there is clearly no relationship across groups for D_{w1} although there may be a curvilinear one for D_{w2}. Regression analyses have been carried out within the $\overline{D} = -2$ group for both D_{w1} and D_{w2} and results are given in Table 3.

TABLE 3

Analysis of Variance for the Regression 5-locus model

a) u(q) on D_{w1}

Source	d.f.	Sum of Square	Mean Square	F.value
Regression	1	0.007	0.007	7.269
Residual	4	0.004	0.001	

b = ~~0.274~~ 0.192, s.e.(b) = ~~0.0311~~ 0.0712, r = ~~0.969~~ 0.863, s.e.(r) = ~~0.110~~ 0.311

σ^2 = ~~0.0004~~ 0.0009

b) u(q) on D_{w2}

Source	d.f.	Sum of Square	Mean Square	F.value
Regression	1	0.009	0.009	16.859
Residual	4	0.002	0.001	

b = 0.0294, s.e.(b) = 0.0072, r = 0.899, s.e.(r) = 0.229

σ^2 = 0.0005

An unbiassed estimate of the error variance, σ^2, can also be obtained from the standard errors of u(q) by pooling (this assumes equality of variance for each u(q) which is not strictly true), let this be

$\bar{\sigma}^2$, in this case $\bar{\sigma}^2 = 0.00022$. A test of fit to the linear regression can then be made by comparing σ^2 and $\bar{\sigma}^2$ with the appropriate F value. For D_{w1}, $\frac{\sigma^2}{\bar{\sigma}^2} = 4.4$ which is significantly different from 1, i.e. the error variance contains more than the sampling error and the regression equation used does not adequately describe the situation. For D_{w2}, $\frac{\sigma^2}{\bar{\sigma}^2} = 2.36$ which is just non significant at the 5% level. In this case the regression analysis was also done using all the data and ignoring the $\bar{D}$ grouping. See Table 4.

TABLE 4.

Analysis of variance for the Regression 5-locus model

u(q) on D_{w2} over all $\bar{D}$ groups.

Source	d.f.	Sum of Sq.	Mean Sq.	F.value
Regression	1	0.0826	0.0826	165.2
Residual	8	0.0040	0.0005	

b = 0.0305 s.e.(b) = 0.0024, r = 0.978, s.e. (r) = 0.005

$\sigma^2 = 0.0005$

In this case $\bar{\sigma}^2 = 0.00015$ and $\frac{\sigma^2}{\bar{\sigma}^2} = 3.33$ so although the regression used is highly significant it does not adequately account for all the variation, from Figure 4.18 it would appear some curvilinear regression would be more appropriate.

To consider the relative merits of D_{w1} and D_{w2} further as predictors of u(q) a six locus model was examined for just one $\bar{D}$ value. Equations and simulation results are given in Table 5,

TABLE 5

No.	Genotype	Equation for D_{w1}	Equation for D_{w2}	c = 0.0250, r = 0.0256: D_{w1}	D_{w2}	u(q) ± s.e.
1	1 1 1 0 0 0 0 0 0 1 1 1	$-3 + 38c - 124c^2 + 168c^3 - 112c^4 + 32c^5$	$\frac{3}{r} + \frac{0}{2r} - \frac{3}{3r} - \frac{2}{4r} - \frac{1}{5r}$	-2.125	1.300	0.835 ± 0.0176
2	1 0 1 1 0 0 0 1 0 0 1 1	$-3 + 14c - 36c^2 + 72c^3 - 80c^4 + 32c^5$	$-\frac{1}{r} - \frac{2}{2r} + \frac{1}{3r} + \frac{0}{4r} - \frac{1}{5r}$	-2.671	-1.867	0.676 ± 0.0151
3	1 1 0 1 0 0 0 0 1 0 1 1	$-3 + 22c - 76c^2 + 136c^3 - 112c^4 + 32c^5$	$-\frac{1}{r} + \frac{0}{2r} + \frac{1}{3r} - \frac{2}{4r} - \frac{1}{5r}$	-2.495	-1.367	0.744 ± 0.0144
4	1 0 0 1 1 0 0 1 1 0 0 1	$-3 + 6c + 4c^2 + 8c^3 - 48c^4 + 32c^5$	$-\frac{1}{r} - \frac{4}{2r} - \frac{1}{3r} + \frac{2}{4r} - \frac{1}{5r}$	-2.847	-2.367	0.660 ± 0.0129
5	1 0 1 0 0 1 0 1 0 1 1 0	$-3 + 6c + 4c^2 - 24c^3 + 48c^4 - 32c^5$	$-\frac{3}{r} + \frac{0}{2r} + \frac{1}{3r} - \frac{2}{4r} + \frac{1}{5r}$	-2.848	-2.967	0.633 ± 0.0117
6	1 0 1 0 1 0 0 1 0 1 0 1	$-3 + 6c - 12c^2 + 40c^3 - 48c^4 + 32c^5$	$-\frac{5}{r} + \frac{4}{2r} - \frac{3}{3r} + \frac{2}{4r} - \frac{1}{5r}$	-2.857	-3.700	0.623 ± 0.0114
7	1 0 0 0 1 1 0 1 1 1 0 0	$-3 + 14c - 4c^2 - 56c^3 + 80c^4 - 32c^5$	$+\frac{1}{r} - \frac{2}{2r} - \frac{3}{3r} + \frac{0}{4r} + \frac{1}{5r}$	-2.653	-0.800	0.729 ± 0.0174

u(q) = the mean chance of fixation in a simulated population of size N = 10, with Nia = 2.5, q = 0.5 for all loci and c = 0.025 between each adjacent pair of loci.

$\overline{D}$ = -3/(4 x 15)

all values and equations for D_{w1} and D_{w2} have been multiplied by 60 to remove the denominator term. D_{w2} had further been multiplied by r. Figure 4.19 gives results togetner with the appropriate regression lines for u(q) on D_{w1} and D_{w2}. Regression analyses are given in Table 6.

TABLE 6

Analysis of Variance for the Regression 6-locus model

(a) u(q) on D_{w1}

Source	d.f.	Sum of sq.	Mean Sq.	F.value
Regression	1	0.032	0.0320	77.31
Residual	5	0.002	0.0004	

b = 0.274, s.e.(b) = 0.0311, r = 0.969, s.e.(r) = 0.110

σ^2 = 0.00040

(b) u(q) on D_{w2}

Source	d.f.	Sum of Sq.	Mean Sq.	F. value
Regression	1	0.032	0.0320	99.47
Residual	5	0.002	0.0003	

b = 0.0447, s.e.(b) = 0.0048, r = 0.975, s.e.(r) = 0.098,

σ^2 = 0.00032

Again the error variance can be estimated from the data in Table 5, this gives $\hat{\sigma}^2$ = 0.00021. For D_{w1} $\frac{\sigma^2}{\hat{\sigma}^2}$ = 1.90, not significantly different from 1. For D_{w2} $\frac{\sigma^2}{\hat{\sigma}^2}$ = 1.52, not significantly different from 1.

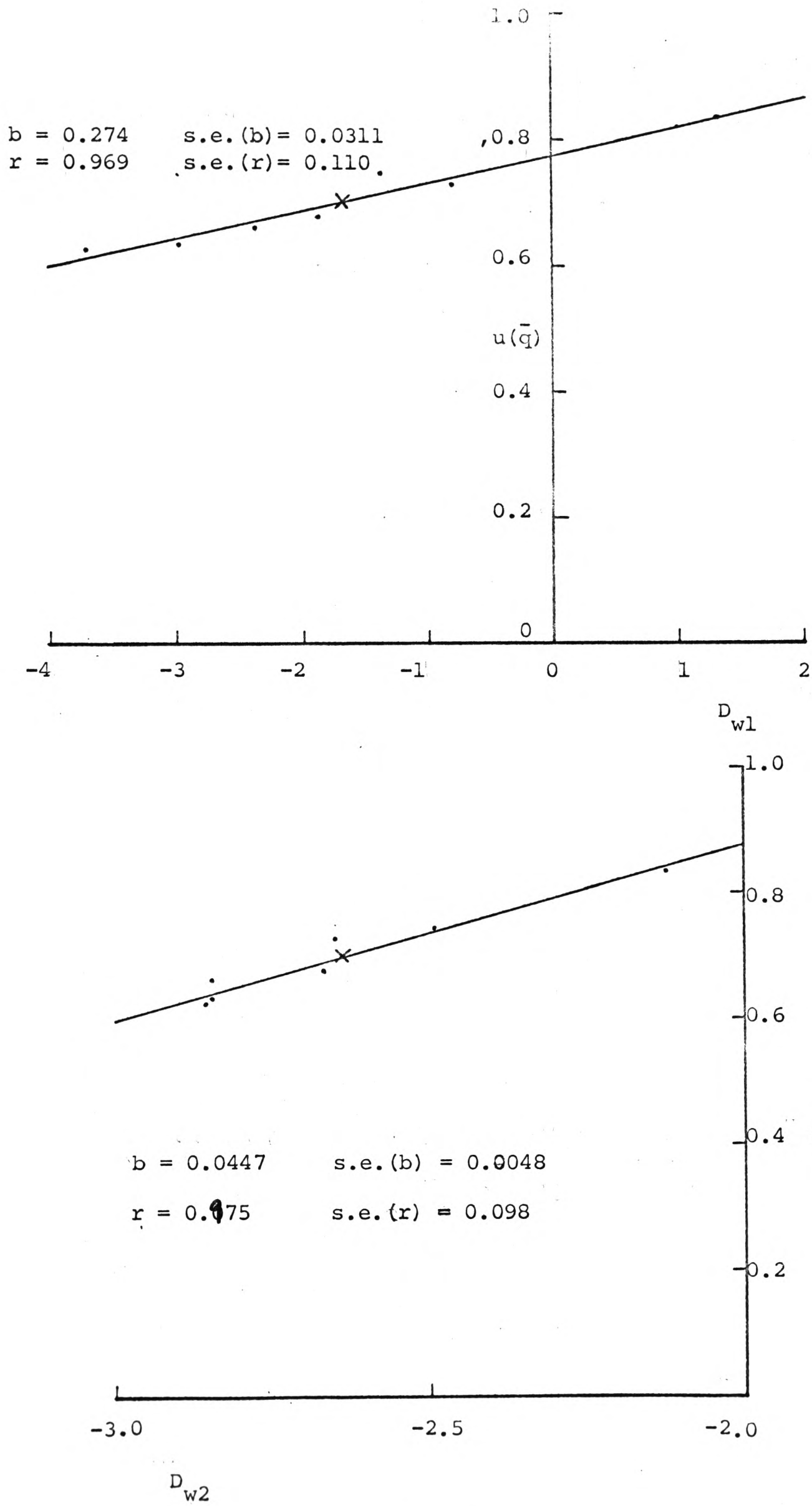


FIGURE 4.19 The regressions of $u(\bar{q})$ on D_{w1} and D_{w2} for N = 10, n = 6, Niα = 2.5, Nc = 0.125. q = 0.5, $D_{w1} = -3/(4^{w2} \times 15)$.

Therefore both predictors adequately account for the differences in u(q) due to differences in order.

From Table 5 the genotypes can be ranked on the basis of their chance of fixation, allowing confidences intervals of plus or minus two standard deviations, this gives

1, 3, 7 (or 7, 3), 2, 4 (or 4, 2) 5, 6 (or 6, 5)

Ranking on the basis of D_{w1} gives

1, 3, 7, 2, 4, 5, 6

Ranking on the basis of D_{w2} gives

1, 7, 3, 2, 4, 5, 6

therefore both agree well with chance of fixation ranking. It is interesting to note here that Fraser et al's (1965) recombination index does not distinguish between genotypes 3 and 4 while on the basis of both D_{w1} and D_{w2} there is a considerable difference. Examination of the difference in u(q) shows a difference of 0.084 which is significant at the 1% level.

From the definitions of D_{w1} and D_{w2} it is apparent that while D_{w2} can be given as a constant term times $^{1}/r$ (if r is the same for each adjacent pair of loci), this is not true of D_{w2} which must contain some term involving the recombination fraction. The significance of this can be seen by considering the comparative weighting of the disequilibrium for pairs separated by r and 2r units, for D_{w2} this is given by r:2r, i.e. the weight given to the near loci is twice that given to the further loci. For D_{w1} it is given by $e^{2r}:e^{4r}$, i.e. the average weight given to the near loci is e^{2r} times that given to the further loci. Therefore while D_2 may only be considered within a particular r value,

D_{wl} can be used over r or c values. Consider for the 6 locus case the 'best' and 'worst' chromosome arrangements, i.e.

(1) $\begin{matrix} 1\,1\,1\,0\,0\,0 \\ 0\,0\,0\,1\,1\,1 \end{matrix}$ and (6) $\begin{matrix} 1\,0\,1\,0\,1\,0 \\ 0\,1\,0\,1\,0\,1 \end{matrix}$

simulation studies showed that chance of fixation for the (6) arrangement was approximately linear with $(c + Nc)/2(1 + 2Nc)$, see Figure 4.20, that is that the value of c tended to exert its influence as a function of $(c + Nc)/(1 + 2Nc)$.

For $c = 0$, $(c + Nc) / (1 + 2Nc) = 0$

$c = 0.5$, $(c + Nc) / (1 + 2Nc) = 0.5$

therefore if c was replaced by $(c + Nc)/(1 + 2Nc)$ in equations for D_{wl}, the cases for $c = 0$ and $c = 0.5$ would be unchanged. Values calculated in this way for arrangements (1) and (6) are given in Table 7 together with the mean chance of fixation obtained by simulation. The appropriate regression analyses are given in Table 8.

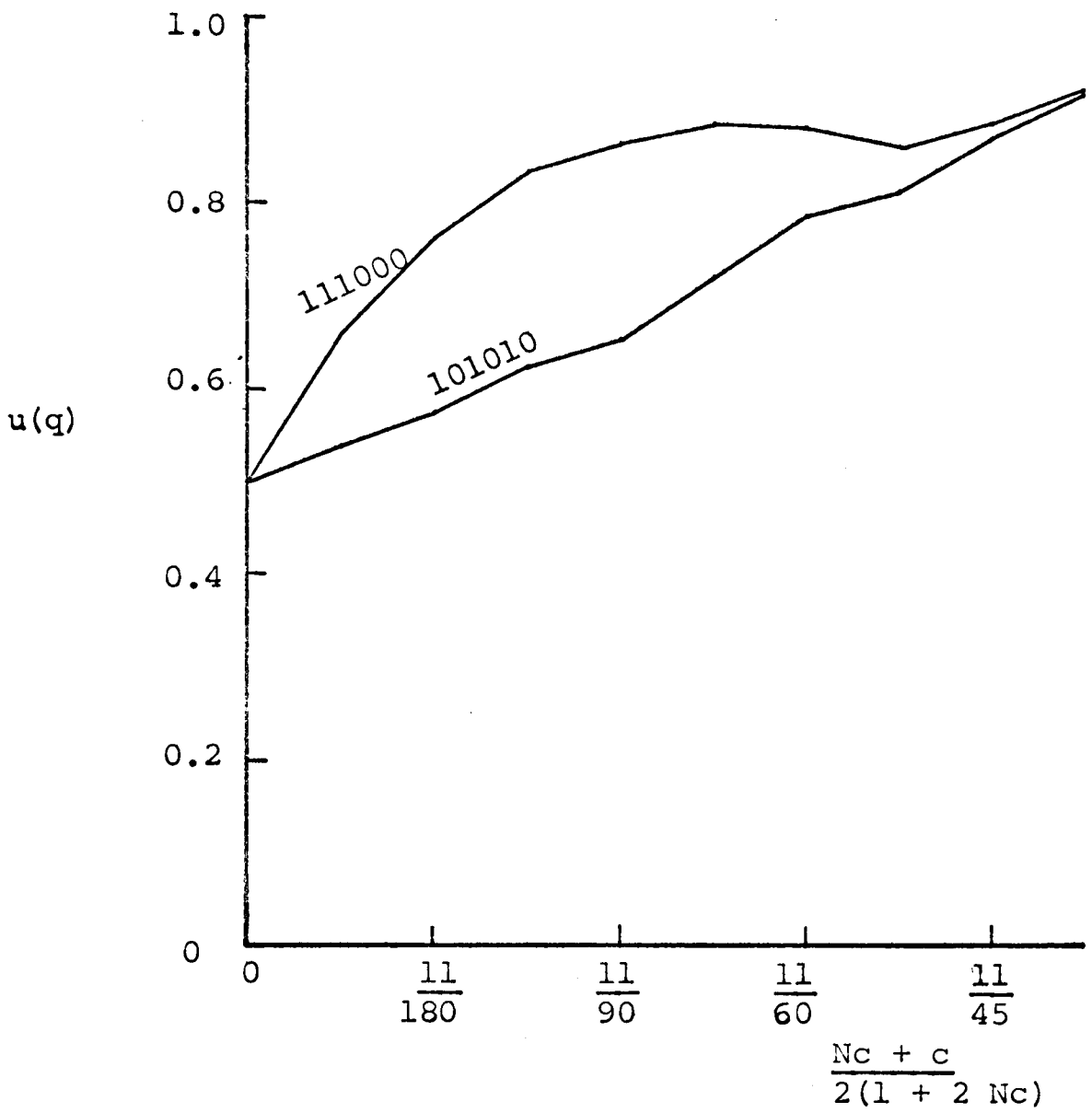


FIGURE 4.20 The relationship between chance of fixation for different chromosome arrangements and recombination fraction plotted as $\frac{c+Nc}{2(1+2Nc)}$ for N = 10, n = 6, Niα = 2.5 q=0.5.

TABLE 7

		1 1 1 0 0 0 0 0 0 1 1 1		1 0 1 0 1 0 0 1 0 1 0 1	
c	$\frac{c + Nc}{1 + 2Nc}$	D_{wl}	u(q) ± s.e.	D_{wl}	u(q) ± s.e.
0	0	-3.00	0.500 -	-3.00	0.500 -
0.00625	0.061	-1.10	0.657 ± 0.0129	-2.67	0.540 ± 0.0110
0.0143	0.122	0.02	0.763 ± 0.0153	-2.37	0.568 ± 0.0122
0.0250	0.183	0.75	0.835 ± 0.0176	-2.12	0.624 ± 0.0114
0.0400	0.244	0.98	0.865 ± 0.0165	-1.80	0.655 ± 0.0123
0.0625	0.305	0.95	0.889 ± 0.0183	-1.49	0.719 ± 0.0117
0.100	0.366	0.75	0.882 ± 0.0175	-1.10	0.787 ± 0.0144
0.17500	0.427	0.43	0.861 ± 0.0149	-0.65	0.806 ± 0.0153
0.5000	0.500	0	0.890 ± 0.0180	0	0.884 ± 0.0172

u(q) = chance of fixation in a simulated population of size N = 10 with equal effects Nia = 2.5 and q = 0.5.

D_{wl} = weighted linkage disequilibrium calculated from equations given in Table 5 with c replaced by (Nc + c)/(1 + 2Nc).

TABLE 8

Analysis of variance for the Regression 6-locus model

(a) u(q) on D_{w1} for arrangement (1)

Source	d.f.	Sum of Sq.	Mean of Sq.	F. value
Regression	1	0.1142	0.1142	27.19
Residual	7	0.0292	0.0042	

b = 0.0924, s.e.(b) = 0.0003, r = 0.892, s.e.(r) = 0.029

σ^2 = 0.0042

(b) u(q) on D_{w2} for arrangement (6)

Source	d.f.	Sum of Sq.	Mean Sq.	F. value
Regression	1	0.1399	0.1399	608.26
Residual	7	0.0016	0.0002	

b = 0.1351, s.e.(b) = 0.00003, r = 0.994, s.e.(r) 0.017

σ^2 = 0.00023

For arrangement (1) $\hat{\sigma}^2$ = 0.00027 ,$\therefore \frac{\sigma^2}{\hat{\sigma}^2}$ = 15.55

while for arrangement (6) $\hat{\sigma}^2$= 0.00017, $\therefore \frac{\sigma^2}{\hat{\sigma}^2}$ = 1.35

so in the former case although the regression is significant it does not adequately remove all the variation due to differences in c.

In conclusion the relative merits of the three systems of measurement can be summarized as follows:

(a) Recombination index.

A simple measure appropriate only for the situation where all individuals are of the same heterozygous genotype. It can

only be used within a single $\bar{D}$ group and even then fails to distinguish between genotypes which have quite large differences in chance of fixation.

(b) D_{w1}.

This represents a weighting system which can be applied to any population where the D_{ij} values are known. It is a parameter combining values of disequilibrium with recombination fraction which can be useful in predicting chance of fixation within a single $\bar{D}$ group.

(c) D_{w2}.

This also represents a weighting system which can be applied to any population where the D_{ij} values are known. It can be defined as

$$D_{w2} = r \sum_{i} \sum_{j} \frac{1}{k_{ij}} D_{ij} \quad \text{where } k_{ij} = \frac{r_{ij}}{r}$$

i.e. such that r appears only as factor. It is in some situations more useful in that it still has a relationship with chance of fixation even across $\bar{D}$ groups.

The relative usefulness of D_{w1} and D_{w2} will depend upon the situations in which they are to be used and on the comparisons which are to be made. In the next section the effect of linkage disequilibrium has been investigated under a multi locus model, mostly for the case where $\bar{D}$ takes its most extreme negative value.

(e) <u>The effect of linkage disequilibrium under the multilocus model</u>.

The importance of linkage disequilibrium has been considered under a model involving four loci, for although this is only a small number it does serve to illustrate the importance

of a fact mentioned earlier, i.e. that for n > 2 for every two pairs of loci in negative disequilibrium there must be at least one pair in positive disequilibrium. With four loci there are six pairs so that a maximum negative $\bar{D}$ is obtained if four are negative disequilibrium while two are in positive disequilibrium. In general for n loci, (i) if n is even, maximum negative disequilibrium is given by $\frac{n(n-2)}{4}$ pairs in positive disequilibrium and $\frac{n^2}{4}$ pairs in negative disequilibrium, (ii) if n is odd; by $\frac{(n-1)^2}{4}$ pairs in positive disequilibrium and $\frac{(n^2-1)}{4}$ pairs in negative disequilibrium. Therefore if the magnitude of the disequilibrium is the same for all pairs $\bar{D}$ is maximized at its most negative for $\bar{D} = \frac{-1}{n-1} D_{ij}$ for n even and $D = -\frac{1}{n} D_{ij}$ for n odd, so that $\bar{D}$ can take less and less extreme negative values as n increases.

The model studied here assumes that linkage disequilibrium was generated by the crossing of two populations each of which have favourable alleles segregating at only two loci, the other two being fixed unfavourably. Each population is assumed to be in linkage equilibrium and the frequency of the favourable allele at each locus is assumed to be the same at $\bar{q}$. The system initially studied had frequencies at the four loci, A,B,C and D as follows:

Frequency of	A	in Population	$1 = 2\bar{q}$
"	A	"	$2 = 0$
"	B	"	$1 = 0$
"	B	"	$2 = 2\bar{q}$
"	C	"	$1 = 2\bar{q}$
"	C	"	$2 = 0$
"	D	"	$1 = 0$
"	D	"	$2 = 2\bar{q}$

Then in the cross the frequencies of chromosomes with two favourable alleles was as follows,

Frequency of chromosomes with A and B = 0

" " A and C = $2\bar{q}^2$

" " A and D = 0

" " B and C = 0

" " B and D = $2\bar{q}^2$

" " C and D = 0

$\therefore$ $D_{AB} = D_{AD} = D_{BC} = D_{CD} = -\bar{q}^2$

$D_{AC} = D_{BD} = \bar{q}^2$

$\therefore$ $\bar{D} = -\bar{q}^2/3$ for $c = 0$ $\bar{D}$ is a sufficient measure of the disequilibrium. As before simulation results for chance of fixation have been compared with those of a similar population which was initially in linkage equilibrium.

The following situations have been considered:

(i) Equal effects at all loci and no recombination.

In this case the chromosome ABCD cannot be formed in the disequilibrium population and the best ones which do occur, i.e. the A-C- and -B-D chromosomes, have only half the value. However even in the equilibrium population the highest initial frequency that the ABCD chromosome can take is $\bar{q}^4$, that is a maximum frequency of 0.0625, so that it may not contribute much to the difference. Chromosomes carrying 3 favourable and one unfavourable allele will also be absent from the disequilibrium population (these will be termed 3f chromosomes, the ABCD being 4f etc), these have a value one and a half times that of the 2f chromosomes present and

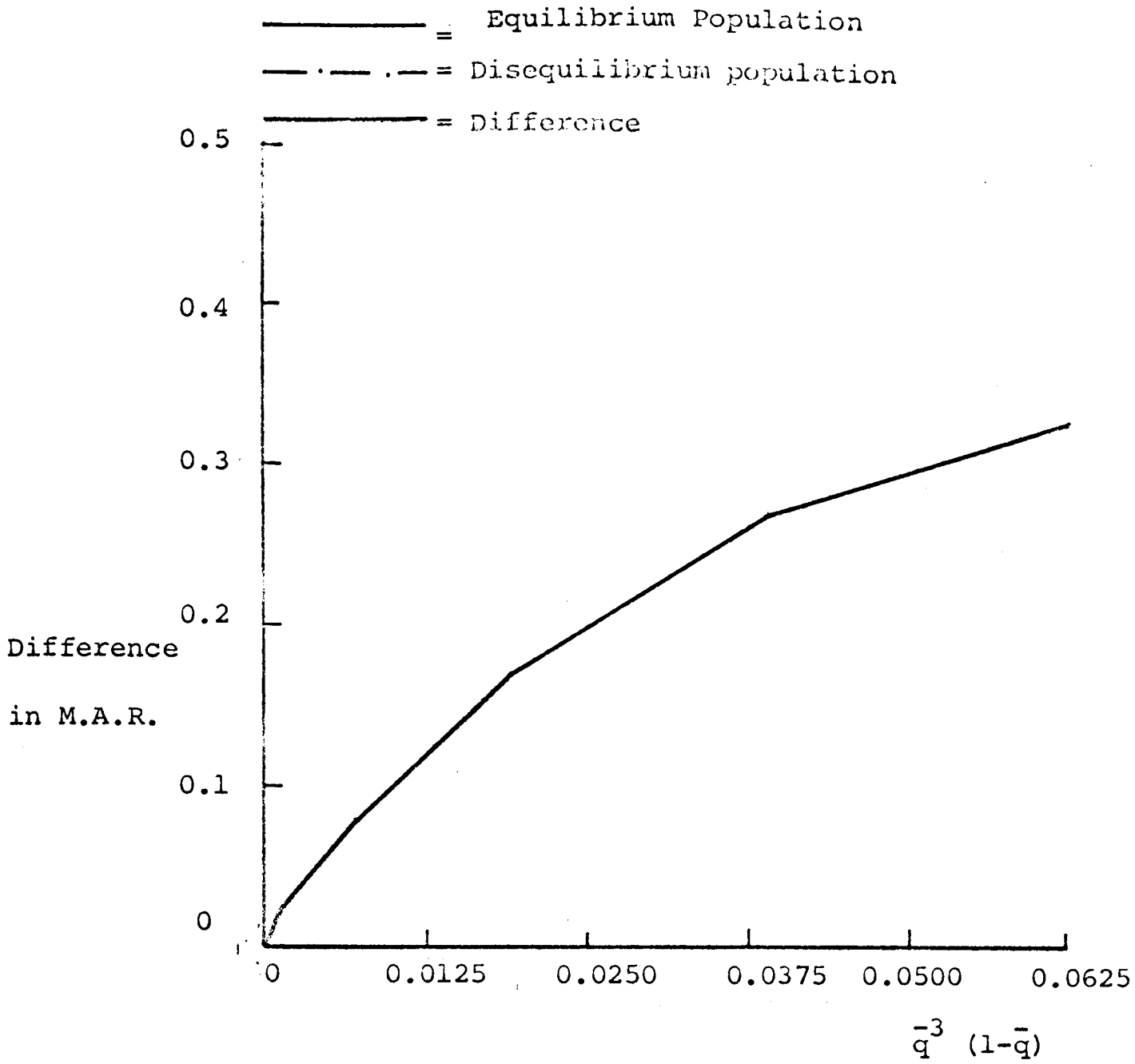
a frequency of $4\bar{q}^3(1-\bar{q})$ which is maximized at 0.25. Therefore it is the fixation of these 3f chromosomes which is likely to contribute most to the difference between the equilibrium and disequilibrium populations.

In the disequilibrium population the highest value chromosomes segregating are the A-C- and -B-D chromosomes each at a frequency of $2\bar{q}^2$, therefore the best which can be done is to fix either one of these given a mean chance of fixation over all loci of $u(\bar{q}) = 0.5$. In the equilibrium population even for $\bar{q} = 0.5$ many of the chromosomes fixed may be of the 3f type so that $u(\bar{q})$ may not greatly exceed 0.75 although obviously this depends on the size of effects. Figure 4.21 shows simulation results for N = 10, $Ni\alpha = 5$, Nc = 0 with the difference in M.A.R. plotted against $\bar{q}^3(1-\bar{q})$ revealing a curvilinear relationship. It is noticeable that the difference in M.A.R. is not as great as was observed for similar parameters for the two locus case, this will be due to the fact that some pairs of loci are in positive linkage disequilibrium and are therefore at an advantage over the equilibrium population,

(ii) Equal effects at all loci with recombination

With recombination all differences between the disequilibrium and equilibrium populations will be reduced.

If an F_1 cross is made and no selection is practised between F_1 individuals then the following chromosome frequencies are expected in the F_2.



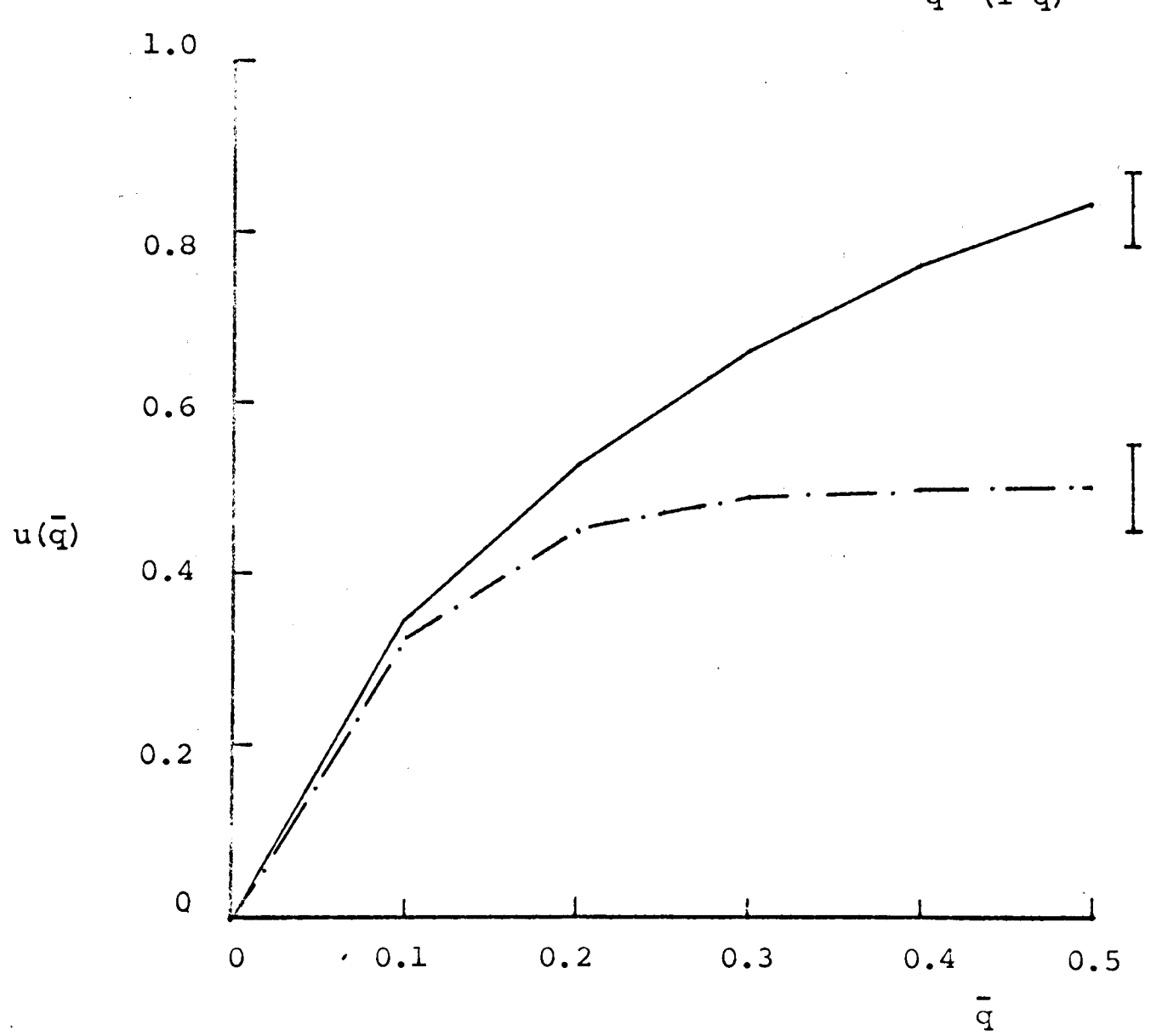


FIGURE 4.21. The effect of initial linkage disequilibrium under a multi-locus model for N=10, n=4, Niα=5, Nc=0 and various values of $\bar{q}$. Typical ranges of length four standard errors are shown.

Frequency of chromosomes carrying A and B = $2\bar{q}^2c$

" " " B and C = $2\bar{q}^2c$

" " " C and D = $2\bar{q}^2c$

" " " A and D = $2\bar{q}^2c(3-6c+4c^2)$

" " " A and C = $2\bar{q}^2(1-2c+2c^2)$

" " " B and D = $2\bar{q}^2(1-2c+2c^2)$

" " " A,B and C = $4\bar{q}^3c^2$

" " " B,C and D = $4\bar{q}^3c^2$

" " " A,C and D = $4\bar{q}^3c(1-2c+2c^2)$

" " " A,B and D = $4\bar{q}^3c(1-2c+2c^2)$

" " " A,B,C and D = $8\bar{q}^4c^3$

which ϕf $c = 0.5$ gives the equilibrium frequencies, otherwise the frequencies of the 3f and 4f chromosomes are always less than in the equilibrium population. The linkage disequilibrium in the F_2 is given by $D_{AB} = D_{BC} = D_{CD} = -\bar{q}^2(1-2c)$, $D_{AD} = -\bar{q}^2(1-2c(3-6c+4c^2))$ $D_{AC} = D_{BD} = \bar{q}^2(1-4c+4c^2)$

i.e. recombination reduces the magnitude of both the positive and negative disequilibria, with recombination $\bar{D}$ no longer fully describes the situation so D_{wl} has been used in this case.

$$D_{wl} = \frac{\bar{q}^2}{3}\ (1-+2c-2c^2+4c^3)$$

e.g. for $c = 0.03125$ $\quad D_{wl} = -0.9393\ \frac{\bar{q}^2}{3}$

An alternative model would be to let

Frequency of A in population 1 = $2\bar{q}$

" A " 2 = 0

" B " 1 = $2\bar{q}$

" B " 2 = 0

cont...

Frequency of C in population 1 = 0

" C " 2 = $2\bar{q}$

" D " 1 = 0

" D " 2 = $2\bar{q}$

then $\bar{D} = \frac{-2}{3}q$ as before but

$$D_{wl} = \frac{-2}{3}q\,(-1+6c-10c^3+4c^3)$$

e.g. for c = 0.03125, $D_{wl} = -0.8221\ \frac{-2}{3}q$

therefore the effect of disequilibrium would be expected to be further reduced under this model. Simulation results for N=10, Niα = 5 and Nc = 0.3125 are shown in Figures 4.22 for the two models. These show firstly how the effect of linkage disequilibrium is reduced quite markedly by only a low value of Nc and secondly that the rearrangement of the loci on the chromosome to give a lower $|D_{wl}|$ greatly reduces the effect of the disequilibrium.

(iii) Unequal effects with no recombination.

In this case the order of ABCD is of no importance with respect to recombination but the arrangement with respect to effects at the loci is important in all cases let

the effect of A = the effect of C = α_1

the effect of B = the effect of D = α_2

then again two situations can be considered.

(a) where

Frequency of A in Population 1 = $2\bar{q}$

" A " 2 = 0

" B " 1 = 0

" B " 2 = $2\bar{q}$

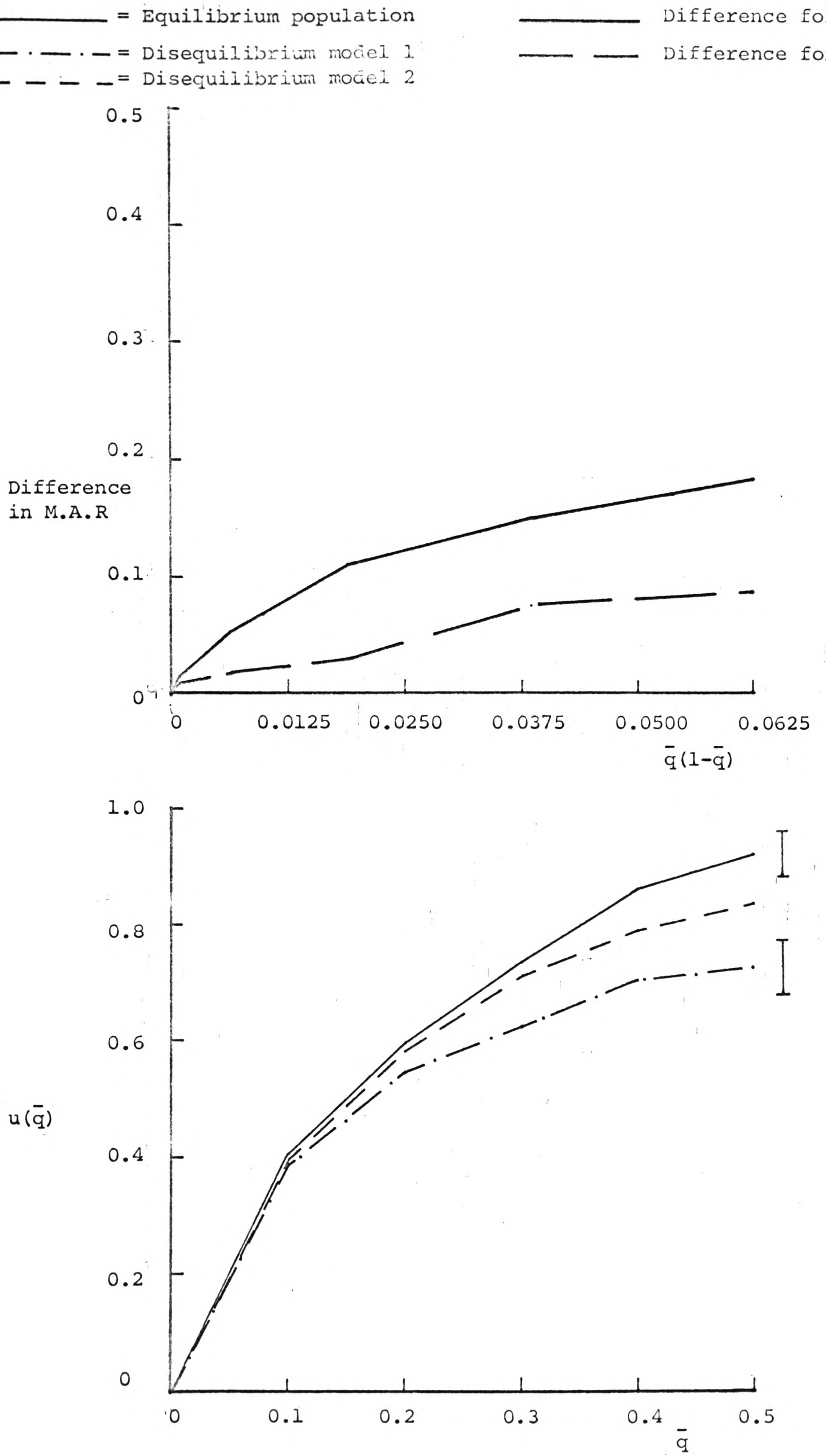


FIGURE 4.22 The effect of initial linkage disequilibrium under a multi-locus model for N=10, n=4, Niα=5, Nc=0.3125 and various values of $\bar{q}$. Typical ranges of length four standard errors are shown.

frequency of C in population 1 = $2\bar{q}$

" C " 2 = 0

" D " 1 = 0

" D " 2 = $2\bar{q}$

Then the mean in population 1 = $4\bar{q}\alpha_1$

and the mean in population 2 = $4\bar{q}\alpha_2$

Let $\alpha_1 > \alpha_2$

If the loci are now considered in pairs there are three types of pairs where

(i) both have effect α_1, i.e. the AC pair

(ii) one has effect α_1, while the other has effect α_2, i.e. the AB, AD, BC and CD pairs.

(iii) both have effect α_2, i.e. the BD pair.

From above $D_{AC} = D_{BD} = +\bar{q}^2$

$D_{AB} = D_{AD} = D_{BC} = D_{CD} = -\bar{q}^2$

that is there is a higher frequency of chromosomes carrying type (i) and type (iii) pairs in the disequilibrium population compared with the equilibrium population, and conversely there is a lower frequency of chromosomes carrying type (ii) pairs. As $\alpha_2 < \alpha_1$, it is a disadvantage for a higher proportion of α_2 type loci to be coupled with other α_2 type loci rather than with α_1 type loci, so that on these grounds alone the chance of fixation of B and D would be expected to be reduced compared with the equilibrium population. Also as $\bar{q}$ increased so the disequilibrium would be increased and therefore the chance of fixation of these loci would be expected to be further reduced. Conversely in the equilibrium population as $\bar{q}$ increased so the chance of fixation of all loci

would be expected to increase. On the other hand it is an advantage for a higher proportion of α_1 type loci to be coupled with other α_1 type loci but this does not necessarily mean that the A and C loci will have a higher chance of fixation in the disequilibrium population, for although the frequency of chromosomes carrying A and C is higher they are all of the type A-C- while in the equilibrium population only $(1-\bar{q})^2$ of them are A-C- with $2\bar{q}(1-\bar{q})$ of the 3f type and $\bar{q}^2$ of the 4f type. Therefore in both populations the chance of fixation of α_1 type loci would be expected to increase with $\bar{q}$. If $\alpha_1 >> \alpha_2$ such that the B and D loci are of little importance then the chance of fixation of A and C would be greater in the disequilibrium population, the difference increasing with $\bar{q}$. If α_1 is only marginally greater than α_2 then the chance of fixation of A and C will be greater in the equilibrium population, the difference increasing with $\bar{q}$.

Figure 4.23 shows simulation results for N=10, $Ni\alpha_1 = 7.5$, $Ni\alpha_2 = 2.5$, Nc = 0. These show chance of fixation of B and D to be greater for the equilibrium population with the difference increasing with $\bar{q}$, while the chance of fixation of A and C is found to be in general greater for the disequilibrium population although differences are always small. As a consequence the difference in M.A.R. is negative for low $\bar{q}$ since the α_1 type loci are of more importance in determining the mean. However as $\bar{q}$ increases the differences in M.A.R. become positive although small.

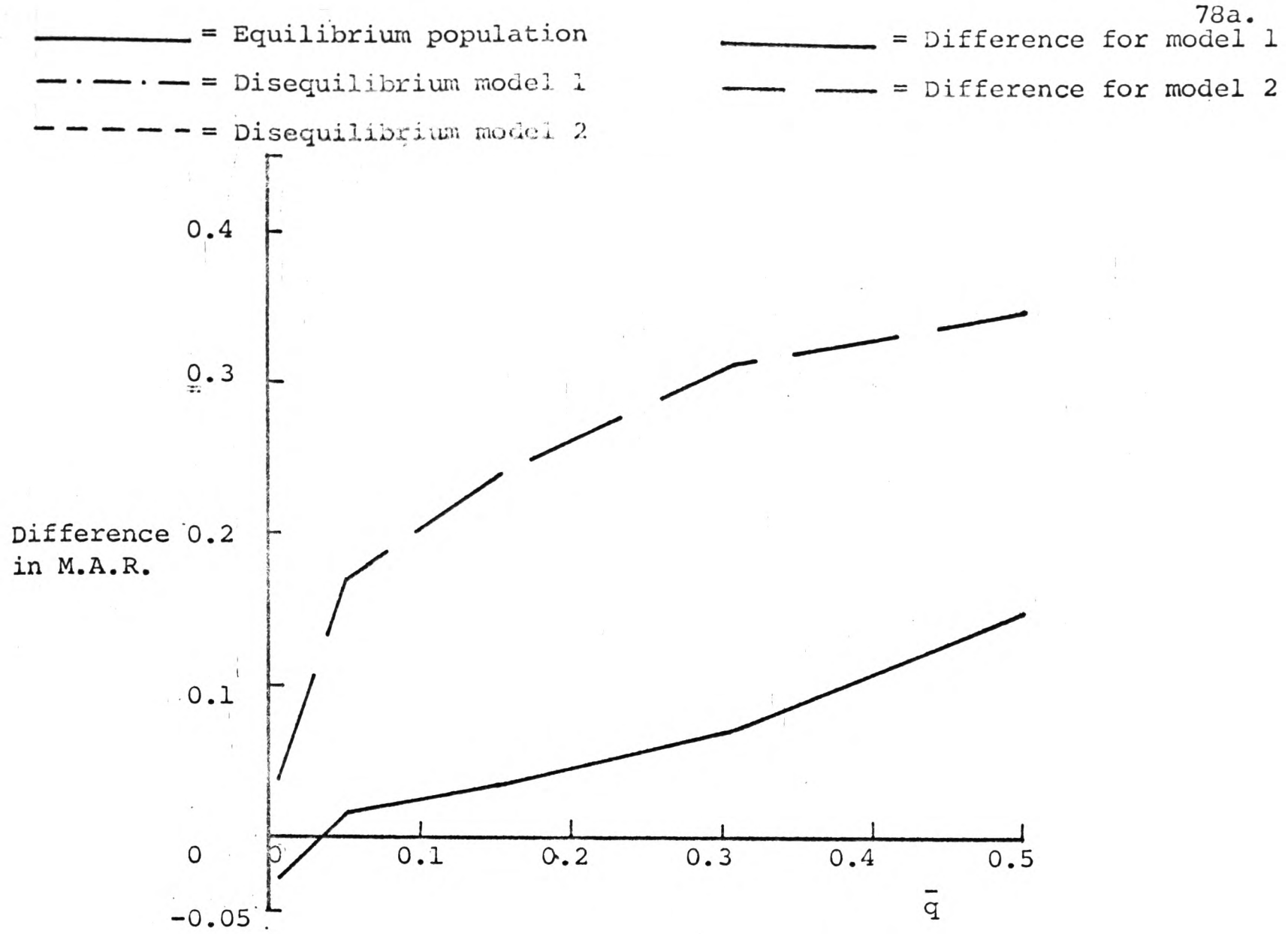


FIGURE 4.23 The effect of initial linkage disequilibrium under a multilocus model with unequal gene effects for N=10, n=4, $Ni\alpha_1$=7.5, $Ni\alpha_2$=2.5; Nc=0 and various values of $\bar{q}$. Typical ranges of length four standard errors are shown.

(b) where

Frequency of A in population 1 = $2\bar{q}$

" A " 2 = 0

" B " 1 = $2\bar{q}$

" B " 2 = 0

" C " 1 = 0

" C " 2 = $2\bar{q}$

" D " 1 = 0

" D " 2 = $2\bar{q}$

Then the mean in population 1 = $2\bar{q}\ (\alpha_1 + \alpha_2)$

and the mean in population 2 = $2\bar{q}\ (\alpha_1 + \alpha_2)$

In this case

$$D_{AB} = D_{CD} = \bar{q}^2$$

$$D_{AC} = D_{AD} = D_{BC} = D_{BD} = \bar{q}^2$$

$\therefore \bar{D} = -\bar{q}^2/3$ as before.

In this case there is a lower frequency of chromosomes carrying type (i) and type (iii) pairs in the disequilibrium population, but there is on average the same frequency of chromosomes carrying type (ii) pairs.

The equilibrium population is the same as for situation (a) since neither gene frequencies nor gene effects have changed. However in the disequilibrium population things are very different since the 2f chromosomes present in this case carry A and B or C and D and are of the same value, therefore the best that can be done is fix one or other of them, giving a maximum chance of fixation to either α_1 type or α_2 type loci of 0.5. Consequently in

comparison with the equilibrium population the chance of fixation of α_1 type loci will be reduced due to three factors:

(a) the absence of 3f and 4f type chromosomes

(b) the reduced frequency of the type (i) pairs

(c) the fact that if A is fixed C must be lost.

On the other hand there are opposing factors affecting the α_2 type loci's chance of fixation,

(a) the absence of 3f and 4f type chromosomes

(b) the reduced frequency of the type (iii) pairs

(c) the fact that if B is fixed D must be lost.

(a) and (c) tend to reduce chance of fixation while (b) tends to increase it. Now (a) is a function of $\bar{q}^3(1-\bar{q})$ and $\bar{q}^4$ while (b) is a function of $\bar{q}^2$, therefore as $\bar{q}$ increases the effect of (a) would be expected to increase faster than the effect of (b) so the chance of fixation of α_2 type loci may be higher in the disequilibrium population but most probably only if $\bar{q}$ is low. Results from simulation are shown together with those under the earlier model in Figure 4.23. As expected differences are large for chance of fixation of α_1 type loci and they increase with $\bar{q}$. For low $\bar{q}$ the chance of fixation of the α_2 type loci is greater in the disequilibrium population but as $\bar{q}$ increases so this difference is reversed. However in spite of this the difference in M.A.R. is at all times positive and is considerably in excess of that under the earlier model, in fact it exceeds that for the equal effects case seen in Figure 21.

(iv) Unequal effects with recombination.

In this case the arrangement of the loci must be considered with

respect to both effects and recombination. Again let

the effect of A = α_1

and the effect of D = α_2

then four possible arrangements have been considered:

Arrangement 1

Chromosome	A	B	C	D
Effects	α_1	α_2	α_1	α_2
Frequency in Pop^n 1	$2\bar{q}$	0	$2\bar{q}$	0
Frequency in Pop^n 2	0	$2\bar{q}$	0	$2\bar{q}$

Arrangement 2

Chromosome	A	B	C	D
Effects	α_1	α_2	α_1	α_2
Frequency in Pop^n 1	$2\bar{q}$	$2\bar{q}$	0	0
Frequency in Pop^n 2	0	0	$2\bar{q}$	$2\bar{q}$

Arrangement 3

Chromosome	A	B	C	D
Effects	α_1	α_1	α_2	α_2
Frequency in Pop^n 1	$2\bar{q}$	0	$2\bar{q}$	0
Frequency in Pop^n 2	0	$2\bar{q}$	0	$2\bar{q}$

Arrangement 4

Chromosome	A	B	C	D
Effects	α_1	α_1	α_2	α_2
Frequency in Pop^n 1	$2\bar{q}$	$2\bar{q}$	0	0
Frequency in Pop^n 2	0	0	$2\bar{q}$	$2\bar{q}$

With respect to the equilibrium population arrangement 1 = arrangement 2 and arrangement 3 = arrangement 4. Under 1 and 2 large loci are more tightly linked to small loci than to other large loci while under 3 and 4 large loci are more tightly linked to other large loci than to small loci. Studies on the effect of linkage per se have shown that it is of greatest importance between loci of equal effect, therefore a higher mean response would be expected for arrangements 1 and 2 compared with 3 and 4. However, simulation results gave no significant difference between the responses in the equilibrium population, this is not really surprising since the change in order has only altered map distances by a factor of two which may make little difference for a small c value, in this case Nc = 0.3125 was used.

In the disequilibrium population various comparisons can be made;

(i) Arrangement 1 with 2 in both these the arrangement of the loci with respect to the effects is the same but under arrangement 1 the important α_1 loci are in positive disequilibrium while under arrangement 2 they are in negative disequilibrium. Therefore it is to be expected that the M.P.R. for 1 will exceed the M.P.R. for 2.

(ii) Arrangement 3 with 4, by the same reasoning it is to be expected that the M.P.R. for 4 will exceed the M.P.R. for 3.

(iii) Arrangement 1 with 4, in both these the positive disequilibrium is between (i) and (iii) type pairs but in arrangement 4 the α_1 type loci are more tightly linked. Therefore it is to be expected that the M.P.R. for 4 will exceed the M.P.R. for 1.

iv) Arrangement 2 with 3, by similar reasoning it is to be expected that the M.P.R. for 2 will exceed the M.P.R. for 3.

Therefore with respect to the M.P.R. the following is to be expected.

Arrangement 1 > Arrangement 2

" 4 > " 3

" 4 > " 1

" 2 > " 3

$\therefore$ 4 > 1 > 2 > 3

Simulation results together with standard errors, for N = 10, Nc = 0.3125, $Ni\alpha_1$ = 7.5, $Ni\alpha_2$ = 2.5, $\bar{q}$ = 0.2 are given in terms of the M.P.R. in Table 12.

TABLE 12

Arrangement	1	2	3	4
Equilibrium	0.5887	0.5887	0.5851	0.5851
	±0.0241	±0.0241	±0.0244	±0.0244
Disequilibrium	0.5216	0.5119	0.4326	0.5564
	±0.0249	±0.0249	±0.0245	±0.0246

This gives the order predicted although in no cases are differences between adjacent values large.

(f) Conclusions

These studies have shown that the negative linkage disequilibrium which may be produced by crossing distinct populations can have a considerable effect on response to selection. The

reduction caused by the disequilibrium tends to have the greatest effect on those loci which have favourable alleles either at low frequency or of small selective advantage.

Linkage disequilibrium can be reduced by a period of relaxation in an infinite population but to have any appreciable effect on the response in the mean at the limit this must be practised for a very long period of time. A further problem concerns size, if a relaxed population is to be of any restricted size the advantages rapidly disappear due to random loss of rare favourable gametes.

The above conclusions have mainly been based on two locus studies, extension to the multi-locus case has shown that as the number of loci, n, increases so the magnitude of the maximum mean negative linkage disequilibrium decreases. Therefore it appears that except for very special circumstances, for example when the two populations in the cross carry different genes of the same effect with no recombination, i.e. virtually a two locus situation, disequilibrium will be of decreasing importance as n increases. Some simulation has been done to investigate the possibility of allowing some generations of relaxation before selecting but no useful advantages have been found suggesting that this procedure is unlikely to be of any practical importance. Practical studies with Drosophila support this idea. Osman and Robertson (1968) crossed selected lines and found no clear advantage in waiting before commencing selection in the cross.

CHAPTER V

Selection Procedures with a Single Base Population

(a) Introduction

This part of the study has been concerned with using population sub-division followed by crossing given a single base population. The work done previously has been reviewed in Chapter II and some of the important theoretical results will be derived below. This work has been mainly for the single locus case and the present study makes some extension of the results already obtained.

For the multilocus case little is known as to the effect of subdivision on limits to selection. In an attempt to clarify this an extensive two locus study has been carried out and the conclusions reviewed in the light of a smaller but rather more generally applicable multilocus study. Because the model used has determined the way in which the problem can be approached the results obtained are best considered separately under the three situations, i.e.

(i) Single locus

(ii) Two locus

(iii) Multi locus

although in all cases the models are intended to be, to some extent at least, representations of the far more complex animal and plant populations.

(b) Single locus models

Let the single locus have alles A/a and let

u(p) = chance of fixation of A in a single population of size N

v(p) = chance of fixation of A under a sub-lining and crossing system with total population size at all times, N.

Robertson (1960) showed that with subdivision into equal lines, selection to the limit, combination and reselection to the limit, u(p) = v(p) for additive loci. This can be seen algebraically using equation (6).For recessive or dominant alleles algebra is not feasible but numerical analysis shows that

v(p) > u(p) for a recessive with p < 0.6 approx.

v(p) < u(p) " " p > 0.6 "

v(p) < u(p) for a dominant with p < 0.6 "

v(p) > u(p) " " p > 0.6 "

In general

v(p) > u(p) for a recessive

v(p) < u(p) for a dominant

However in no cases are differences very large, the maximum found over a range of $Ni\alpha$ and p values was < 0.1 for $Ni\alpha = 10$, $p = 0.1$.

Maruyama (1970) generalized this result for the additive case considering instead of separate selected lines, partially isolated colonies with selection and migration. He showed that providing migration does not change the total number of A alleles in the population the chance of fixation remains the same as if

the population had not been subdivided. For the additive case this result predicts that if a population of size N is subdivided into two lines of unequal size, say hN and (1-h)N ($h = \frac{1}{N}, \frac{2}{N} \ldots \frac{N-1}{N}$) and these are selected to the limit, then crossed and reselected to the limit in a single line size N, the chance of fixation will equal u(p), providing that the crossing process does not alter the total number of A alleles present, i.e. providing the numbers of individuals used from each line in the cross are in proportion to the size of the line. Then the gene frequency in the cross may be 0, 1, h or (1-h). Now using

$$u(p) = \frac{1-e^{-2Ni\alpha p}}{1-e^{-2Ni\alpha}}$$

let

$u(p) = u_1$ = the chance of fixation in the subline of size hN

$$= \frac{1-e^{-2Ni\alpha hp}}{1-e^{-2Ni\alpha h}}$$

$u_2(p) = u_2$ = the chance of fixation in the subline of size (1-h)N

$$= \frac{1-e^{-2Ni\alpha(1-h)p}}{1-e^{-2Ni\alpha(1-h)}}$$

$u(h) = u_3$ = the chance of fixation in the cross if the frequency of A is h

$$= \frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}}$$

$u(1-h) = u_4$ = the chance of fixation in the cross of the frequency of A is (1-h)

$$= \frac{1-e^{-2Ni\alpha(1-h)}}{1-e^{-2Ni\alpha}}$$

Then

$$v(p) = u_1u_2 + u_1(1-u_2)u_3 + u_2(1-u_1)u_4$$

$$= \frac{1-e^{-2Ni\alpha p}}{1-e^{-2Ni\alpha}} \qquad = u(p)$$

Consider now the effect of a different method of crossing, suppose the subline of size N contributes a proportion of y to the cross ($y = \frac{1}{N}, \frac{2}{N} \ldots \frac{N-1}{N}$) let the chance of fixation in this case be given by $v_g(p)$ and let

$u(y) = u_6$ = the chance of fixation in the cross of the frequency of A is y

$$= \frac{1-e^{-2Ni\alpha y}}{1-e^{-2Ni\alpha}}$$

($u(1-y) = u_6$ = the chance of fixation in the cross if the frequency of A is (1-y)

$$= \frac{1-e^{-2Ni\alpha(1-y)}}{1-e^{-2Ni\alpha}}$$

$$\therefore v_y(p) = u_1u_2 + u_1(1-u_2)u_5 + u_2(1-u_1)u_6$$

Let $D_y = v_y(p) - u(p) = v_g(p) - v(p)$

$$= u_1(1-u_2)(u_5-u_3) + u_2(1-u_1)(u_6-u_4)$$

$$= \frac{1-e^{-2Ni\alpha(y-h)}}{1-e^{-2Ni\alpha}} \left[e^{-2Ni\alpha h} u_1(1-u_2) - e^{-2Ni\alpha(1-y)} u_2(1-u_1) \right] \quad \ldots(51)$$

if $y = h$, $D_y = 0$.

A particular case of interest is for y = 0.5, this might occur in practice if one line contributed males and the other females to the cross, then

$$D_y = \frac{1-e^{-Ni\alpha(1-2h)}}{1-e^{-2Ni\alpha}}\left[e^{-2Ni\alpha h}\, u_1(1-u_2) - e^{-Ni\alpha}\, u_2(1-u_1)\right]$$

If $h = 0.5$ $u_1 = u_2$ and $D_y = 0$

If $p = 0.5$ $\dfrac{u_1(1-u_2)}{u_2(1-u_1)} = \dfrac{e^{-Ni\alpha(1-h)}}{e^{-Ni\alpha h}} = \dfrac{e^{-Ni\alpha}}{e^{-2Ni\alpha h}}$. $\therefore \dfrac{D_y}{dh} = 0$

Differentiating D_y w.r.t. h and putting h = 0.5 gives

$$\frac{d\,D_y}{dh} = 0$$

Differentiating a second time w.r.t. h and putting h = 0.5 gives

$$\frac{d^2D_y}{dh^2} = \frac{8\,N^2i\alpha^2\, e^{-Ni\alpha}\, e^{-Ni\alpha p}}{(1-e^{-2Ni\alpha})(1-e^{-Ni\alpha})^2}\left[(1-2p)(1-e^{-Ni\alpha}) + e^{-Ni\alpha(1-p)} - e^{-Ni\alpha p}\right]$$

if $p = 0$ $\dfrac{d^2D_y}{dh^2} = 0$

if $p = 0.5$ $\dfrac{d^2D_y}{dh^2} = 0$

if $p = 1$ $\dfrac{d^2D_y}{dh^2} = 0$

For positive values of Niα it can be shown that

$\frac{d^2D_y}{dh^2}$ is positive for $0 < p < \frac{1}{2}$

then $h = \frac{1}{2}$ gives a minimum value of D_y

$\therefore$ as h moves away from 0.5 the value of D_y increases from zero

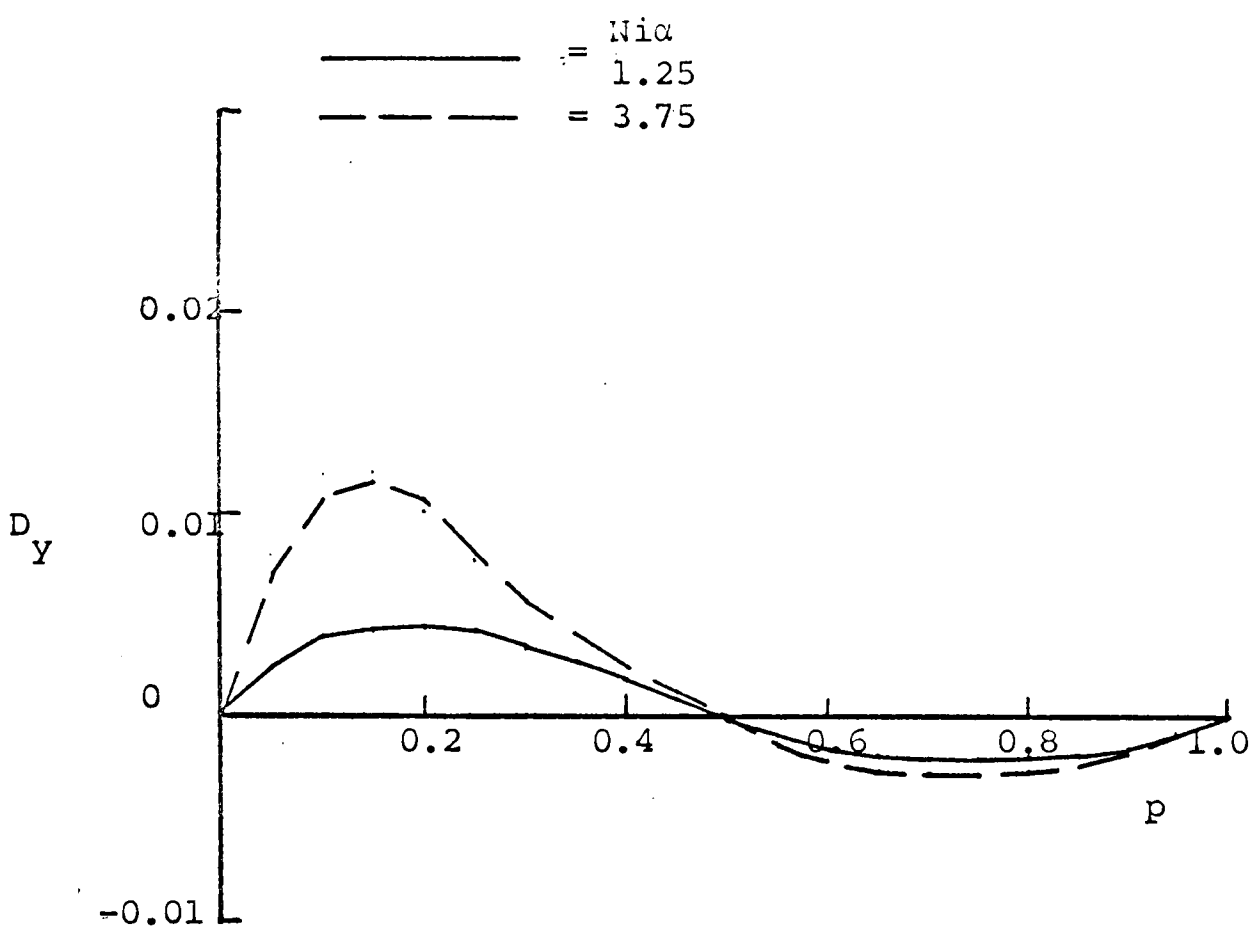
$\frac{d^2D_y}{dh^2}$ is negative for $\frac{1}{2} < p < 1$

then $h = 0.5$ gives a maximum of D_y and as h moves away from 0.5 the value of D_y decreases from zero.

The way in which D_y varies with p and $Ni\alpha$ is shown in Figure 5.1 for $N = 10$, $h = 0.2$. Several other values of N have also been studied for $h = \frac{1}{N}$ with $N = 10$, 20, 40 and 100. These show that in all cases the absolute value of D_y at its maximum positive value is considerably greater than at its maximum negative value. Thus if a high chance of fixation is the aim then the sublining and equal crossing system will on average be superior to single line selection. Highest values of D_y are obtained for $Ni\alpha$ values between 2 and 5, however under no conditions is D_y large, even for $N = 100$, $h = 0.01$ the maximum value of D_y is less than 0.035.

Taking the case $y = 0.5$ was somewhat arbitrary, it being a value which might occur in practice, it is possible that a value of y could be used to further increase D_y. This is given by differentiating D_y w.r.t. y and setting the differential to zero, this gives

$$y_{max} = \log_e \left(\frac{u_1(1-u_2)}{u_2(1-u_1)} \right) / 4\,Ni\alpha + \frac{1}{2} \qquad \ldots(52)$$



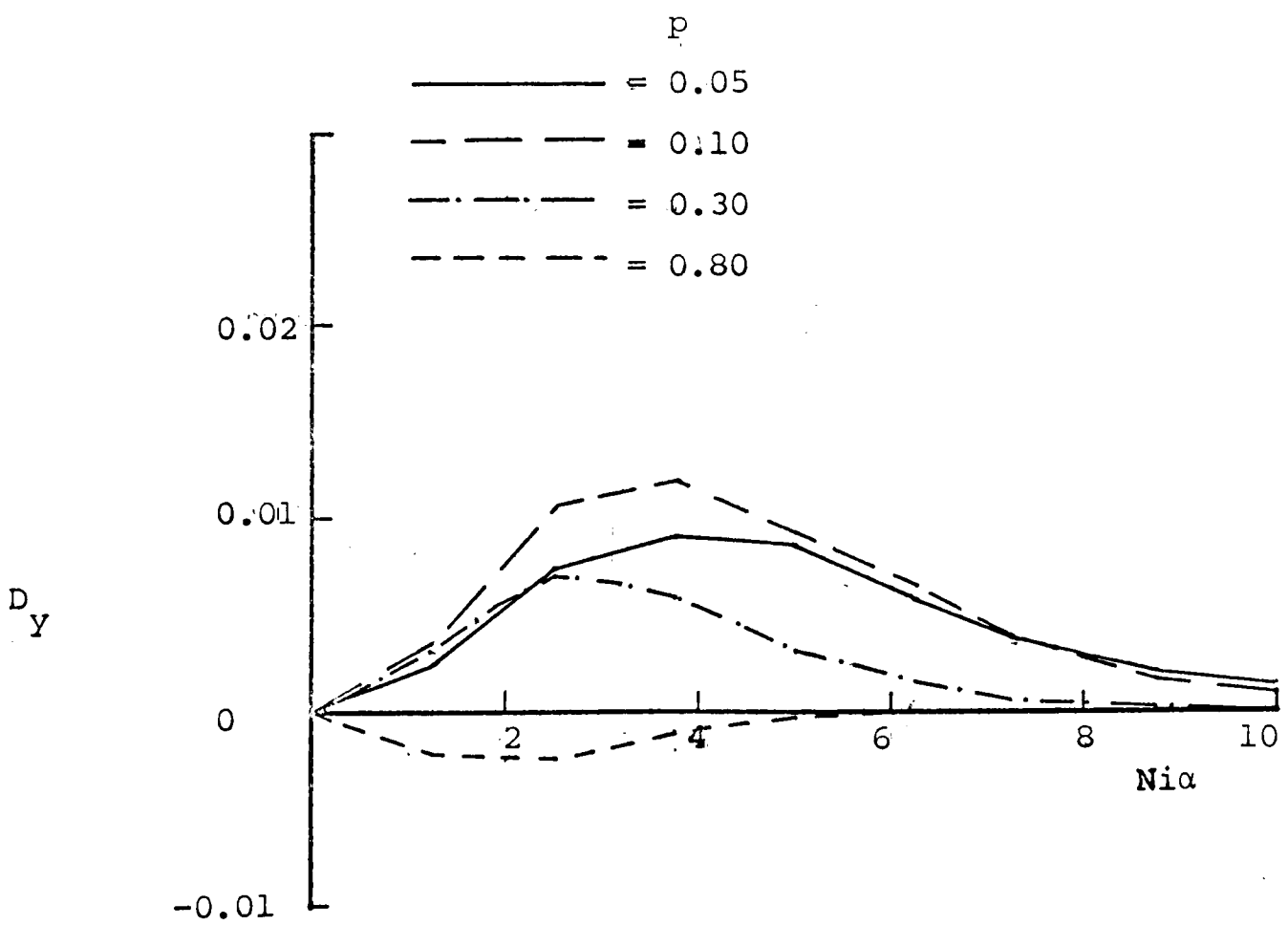


FIGURE 5.1. The relationship between D_y and p and Niα for N=10, n=1, h=0.2, y=0.5 under an additive model.

y_{max} is the value of y which maximizes D_y (the second differential w.r.t. y being negative for all positive Niα).

If h = 0.5, y_{max} = 0.5, therefore for sublines of equal size the greatest limit is attained by either crossing equally, or alternatively not sub-lining at all.

If $\rho = 0.5$, $y_{max} = \frac{h}{2} + \frac{1}{4}$

and $$D_{y_{max}} = \frac{e^{-Ni\alpha(1+h)} \left[1-e^{-Ni\alpha(\frac{1}{2}-h)}\right]^2}{(1-e^{-2Ni\alpha})\ (1+e^{-Ni\alpha(1-h)})\ (1+e^{-Ni\alpha h})}$$

if $h = 0.5$ $D_{y_{max}} = 0$, if $h \neq 0.5$ $D_{y_{max}} > 0$.

For $h < 0.5$ and $p \neq 0.5$, $y_{max} < 0.5$ as Niα increases so y_{max} approaches 0.5. For a given Niα as h moves away from 0.5 so y_{max} moves away from 0.5.

If $y = y_{max}$ $\quad \frac{u_1(1-u_2)}{u_2(1-u_1)} = e^{-2Ni\alpha(1-2y)}$

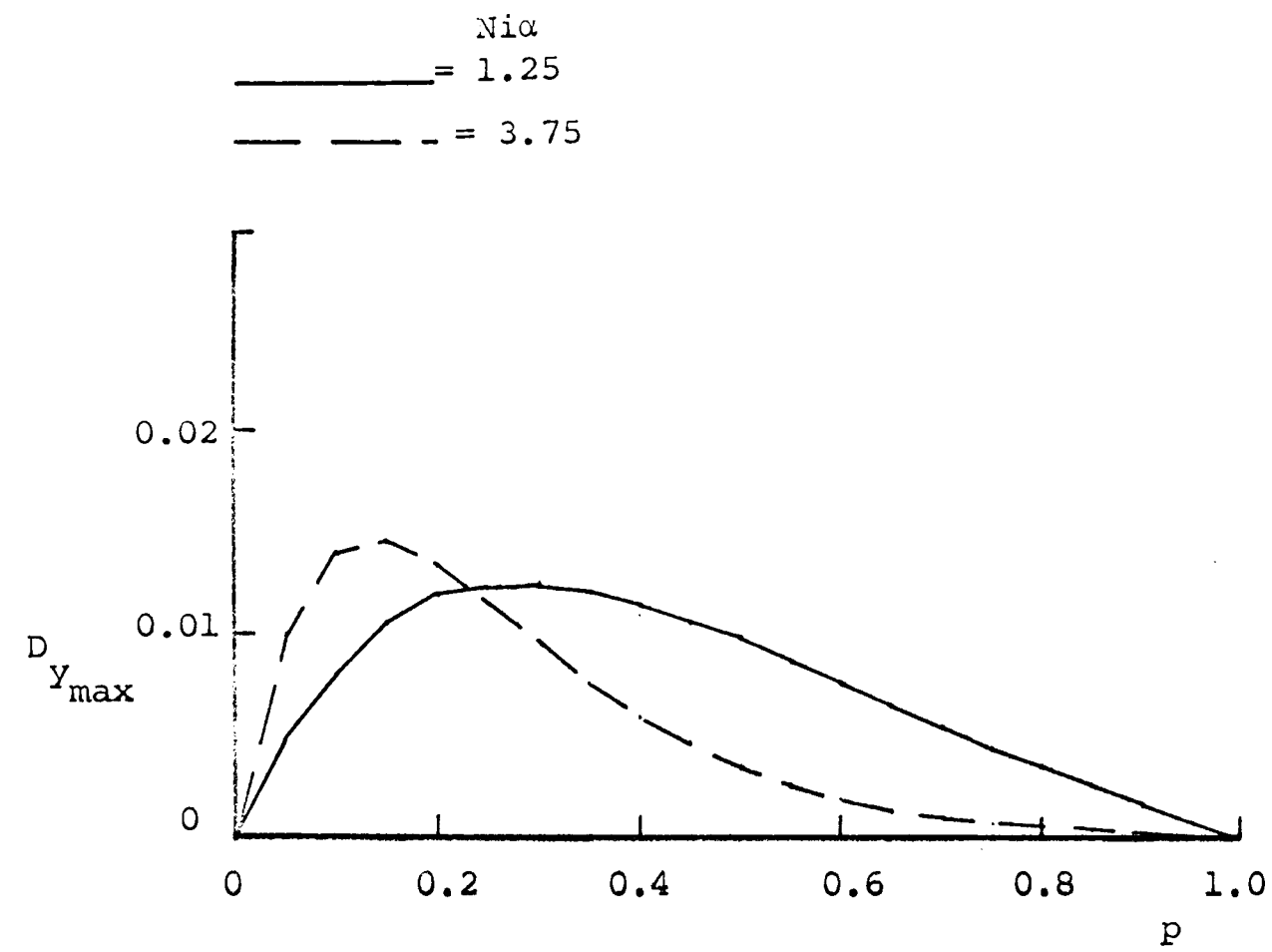
$$\therefore D_{y_{max}} = \frac{\left[1-e^{-2Ni\alpha(y-h)}\right]^2 \quad u_2(1-u_1)\ e^{-2Ni\alpha(h+1-2y)}}{(1 - e^{-2Ni\alpha})}$$

therefore for all positive values of Niα $D_y \geq 0$ of $y = y_{max}$ and for $h \neq 0.5$, $0 < p < 1$ $D_y > 0$.

The relationship of $D_{y_{max}}$ with p and Niα is shown for N = 10, h = 0.2 in Figure 5.2 again emphasising that in no cases are the gains due to this system anything but small.

Dominance and Recessive Models

Under these models of gene action three selection systems have been compared using numerical analysis:



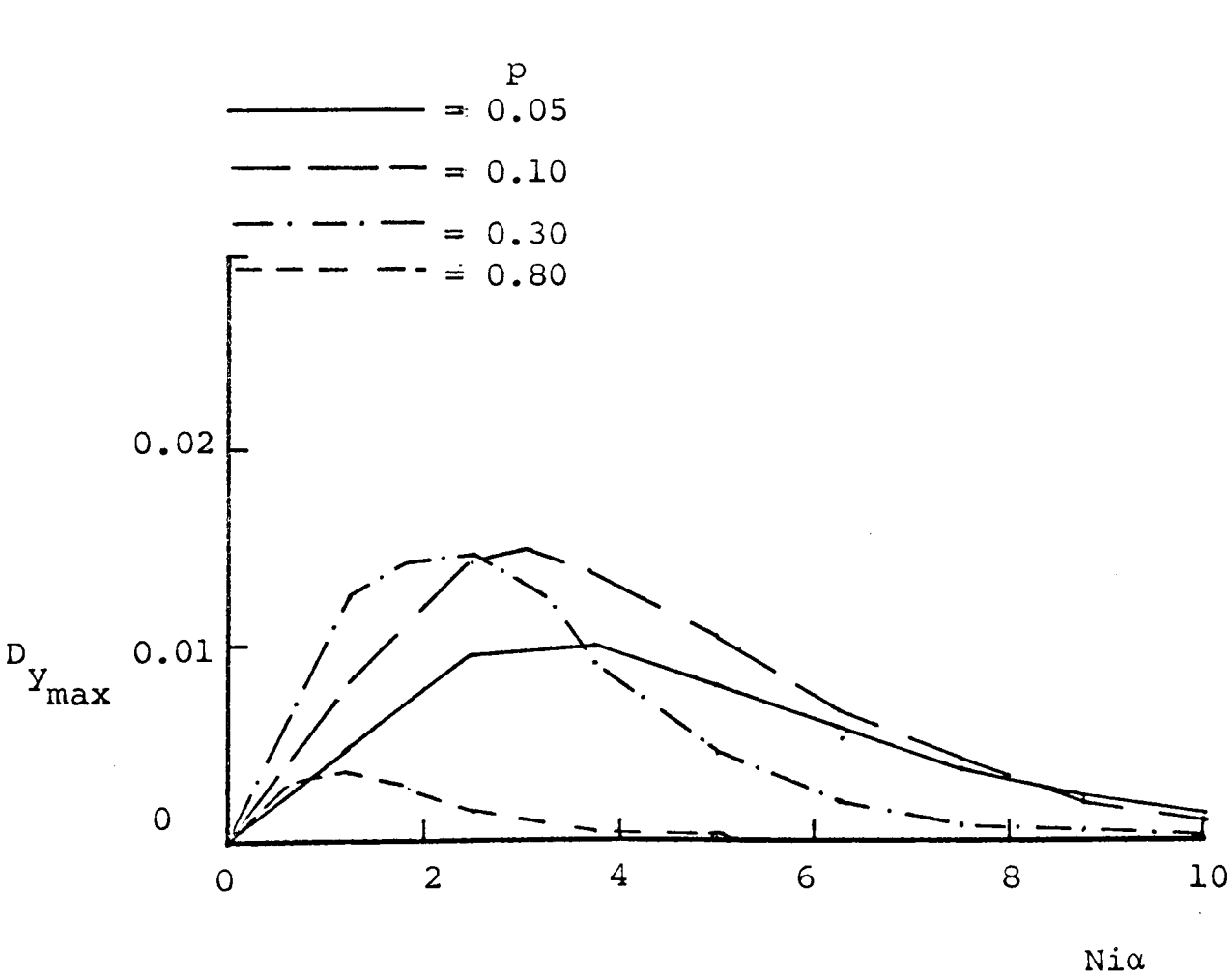


FIGURE 5.2. The relationship between D and p and Niα for n=10, n=1, h=0.2, $y=y_{max}$ where

$$y_{max} = \log_e\left(\frac{u_1(1-u_2)}{u_2(1-u_1)}\right) / 4\ Ni\alpha + 0.5$$

under an additive model.

i) Selection in a single population of size N, chance of fixation = u(p)

ii) Selection to fixation in sublines of sizes hN and (1-h)N with proportional crossing, i.e. y = h, ultimate chance of fixation = v(p)

iii) Selection again in sublines of sizes hN and (1-h)N but with y = 0.5. Ultimate chance of fixation $v_y(p)$.

If h = 0.5 $v_y(p) = v(p)$ but in general

$v(p) \neq u(p)$

Consider:

(a) Selection for a recessive

for h = 0.5 $v(p) > u(p)$ for $p < 0.6$ approx

$v(p) < u(p)$ for $p > 0.6$ "

for $h \neq 0.5$

As h moves away from 0.5 towards 1/N or (N-1)/N so

$|v(p) - u(p)|$ decreases for all p

$|v_y(p) - u(p)|$ decreases for $p < 0.6$ approx

but increases for $p > 0.6$ "

Therefore sub-division of the population into sublines of equal size gives on average the highest limit.

(b) Selection for a dominant.

for h = 0.5

$v(p) < u(p)$ for $p < 0.6$ approx

$v(p) > u(p)$ for $p > 0.6$ "

for $h \neq 0.5$

As h moves away from 0.5 towards 1/N or (N-1)/N so

$| v(p) - u(p) |$ decreases for all p

$| v_y(p) - u(p) |$ decreases for all p

Therefore selection in a single population of size N gives on average the highest limit.

The results obtained for the different sub-lining systems so far examined are summarized in Table 9. The effects of the systems are given in terms of gains and losses relative to the single population case. Typical values of the numerical difference in v (p) - u(p) are given in brackets for N = 10, ia = 0.375 to indicate the order of magnitude of the differences involved.

Consider now the case where the population is split into sublines only for a fixed number of generations and not necessarily to fixation. For the additive case with proportional crossing Maruyama (1970) showed that v(p) is still equal to u(p). To examine some of the other sub-lining systems under this restriction matrix techniques have been used. A square matrix of dimension 2N + 1 was set up with elements b_{ij} where i, j = 0, 1, 2 ... 2N and

$$b_{ij} = \binom{2N}{j} (p + \delta p)^j (1-(p + \delta p))^{2N-j} \quad \ldots(54)$$

where $p = \frac{i}{2N}$ δp = the expected change in gene frequency due to selection.

TABLE 9

	ADDITIVE			DOMINANT		RECESSIVE	
	p < 0.5	p = 0.5	p > 0.5	p < 0.6	p > 0.6	p < 0.6	p > 0.6
1. Effect of 2 equal sized sublines crossed equally	None (0)	None (0)	None (0)	Small losses (-0.03)	v.v.small gains (+0.0003)	Small gains (+0.08)	v.small losses (-0.001)
2. Effect of 2 unequal sublines crossed in proportion to their size.	None (0)	None (0)	None (0)	Losses reduced from (1) (-0.015)	Gains reduced from (1) (+0.0002)	Gains reduced from (1) (+0.035)	Losses reduced from (1) (-0.0008)
3. Effect of 2 unequal sublines crossed using equal numbers	Small gains (+0.01)	None (0)	v.small losses (-0.001)	Losses reduced from (1) (-0.013)	Gains reduced from (1) (+0.0001)	Gains slightly reduced from (1) (+0.07)	Losses increased from (1) (-0.002)
4. Effect of 2 unequal sublines crossed with maximizing y value	Gains increased from (3) (+0.013)	v. small gains (+0.003)	v.v.small gains (+0.0001)	-	-	-	-

Table 9. The effects of different sub-lining systems compared to single line selection for different modes of gene action and frequency.

Repeated iteration of this matrix onto a vector of initial gene frequencies gives the probability that the population contains j A alleles at any given generation. Results for u(p) are obtained in this way as the probability that the population contains 2N A alleles after very many generations, values are not identical with those obtained using equations 6, 7 and 8 but except for large N and low p values differences are small. Using this matrix iteration system sublines can be set up and selected for T generations and then crossed and reselected to fixation. Let,

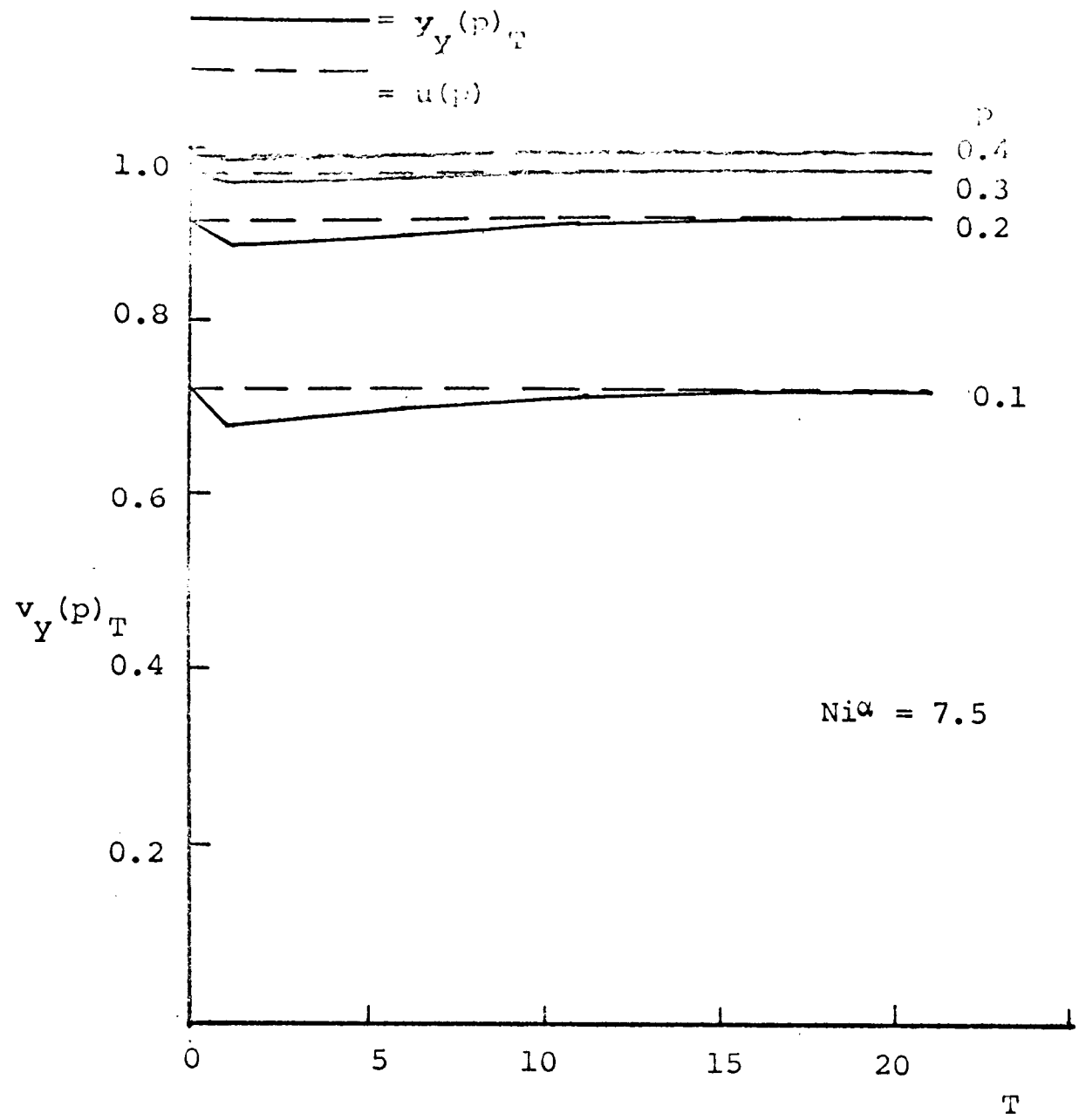
$v(p)_T$ = the chance of fixation if a sub-lining system is used for T generations.

$v(p)_T$ for an additive locus has been found to be independent of T and equal to u(p), this is in agreement with Maruyama's result.

However for y = 0.5, $v_y(p)_T$ is no longer independent of T unless h = 0.5 when y = h as above, and it may be greater or less than u(p) depending on the value of T.

The way in which $v_y(p)_T$ varies with T for an additive model is shown in Figure 5.3 for N = 10, h = 0.2, $i\alpha$ = 0.25, 0.75 for a range of p values. In this case the difference $v_y(p)_T - u(p)$ for T = 20 is approximately equal to D_y but for T = 10 the difference is approximately zero being negative for $0 < T < 10$.

A similar effect has also been found for the case of selection for a recessive where it was found that $v(p) > u(p)$ if $p < 0.6$. In this case $v(p)_T$ is not independent of T even



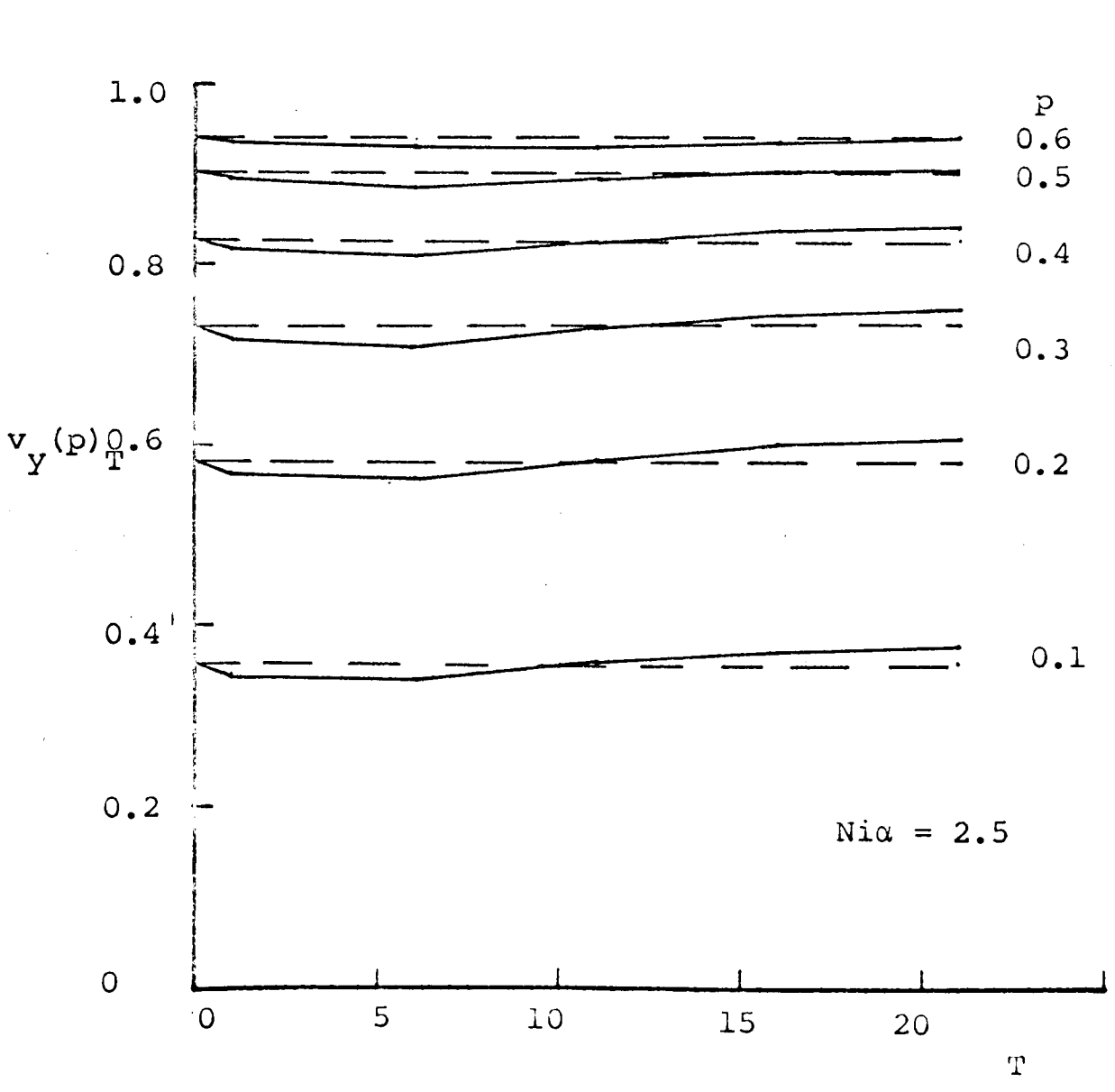


FIGURE 5.3. The relationship between chance of fixation under a sub-lining system and the number of generations of sub-lining, for N=10, n=1, h=0.2, y=0.5, for various values of p and Niα, under an additive model.

for h = 0.5, but it always exceeds u(p) for T > 0. Some results for N = 10, $i\alpha$ = 0.25, 0.75 are shown in Figure 5.4. Here v(p) reaches its maximum value for T ≈ 15. These have simply been examples of the way in which T affects the relationships between the u(p) and v(p) values, the particular value of T at which the sub-lining system would give its maximum limit would be expected to depend on the value of N.

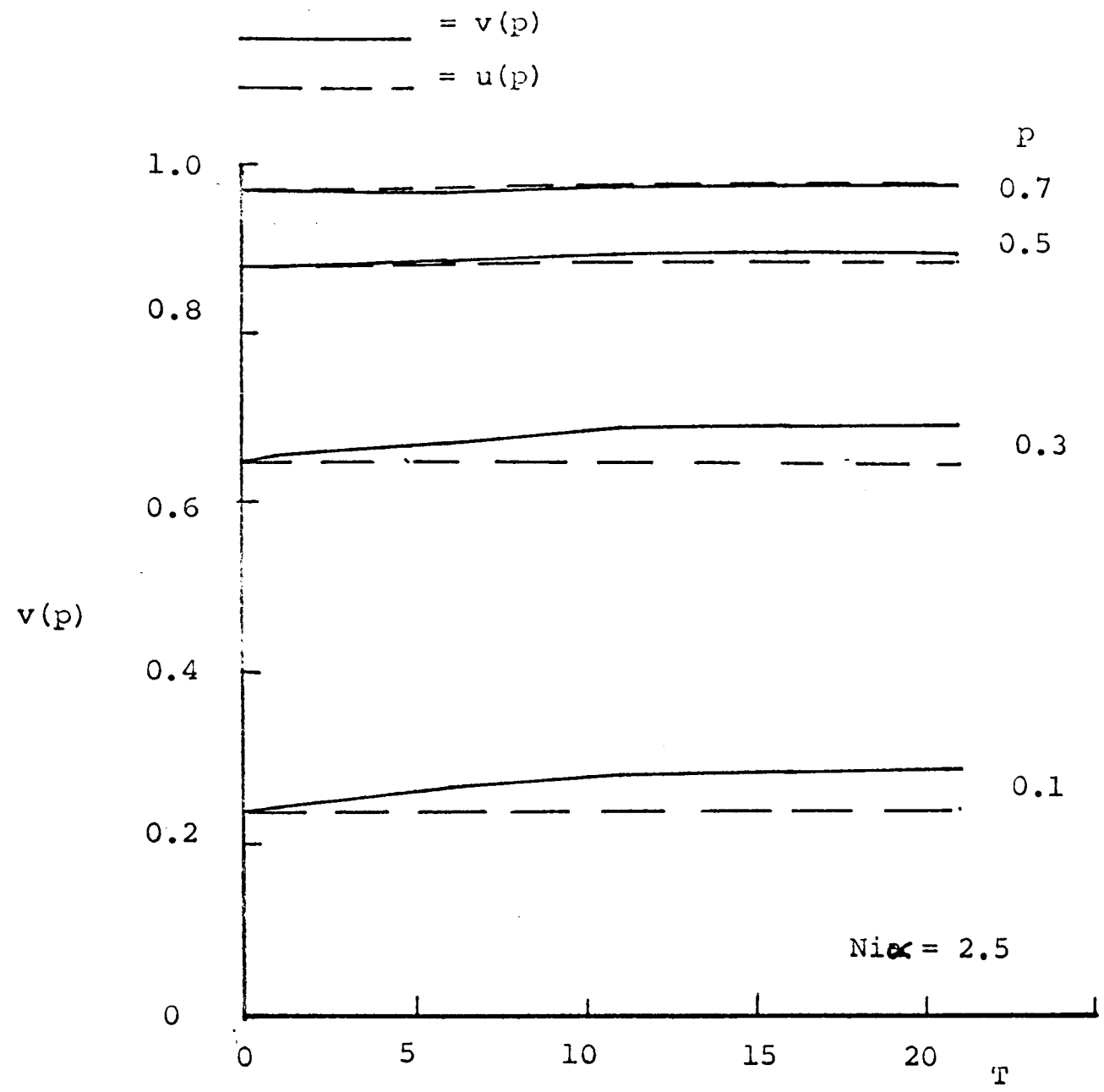
These results have shown that some systems of sub-lining and crossing can, in some cases give a higher chance of fixation than equivalent single line selection. Two cases in particular have given the most advantage, these are:

i) for additive loci, unequal sub-lines selected to fixation and crossed in proportion to y_{max}. Gains are of the order of 1% and there are several problems.

(a) Sub-lining must be to fixation or at least for many generations; if it is for only a short period losses in ultimate chance of fixation may be incurred. This then drastically reduces the overall rate of response compared with the single line system.

(b) y_{max} is a function of p and Niα which will not in general be known. In practice the best that can be done is to put y = 0.5 which further reduces the gains available.

(c) If p is initially high and y = 0.5 is used the chance of fixation may be reduced. A question which has not really been considered so far is why should a value of y not equal to h tend to give a higher limit in some situations? Some insight into this can be gained by considering the case of a gene of quite high selective advantage initially at low frequency. If two sub-lines



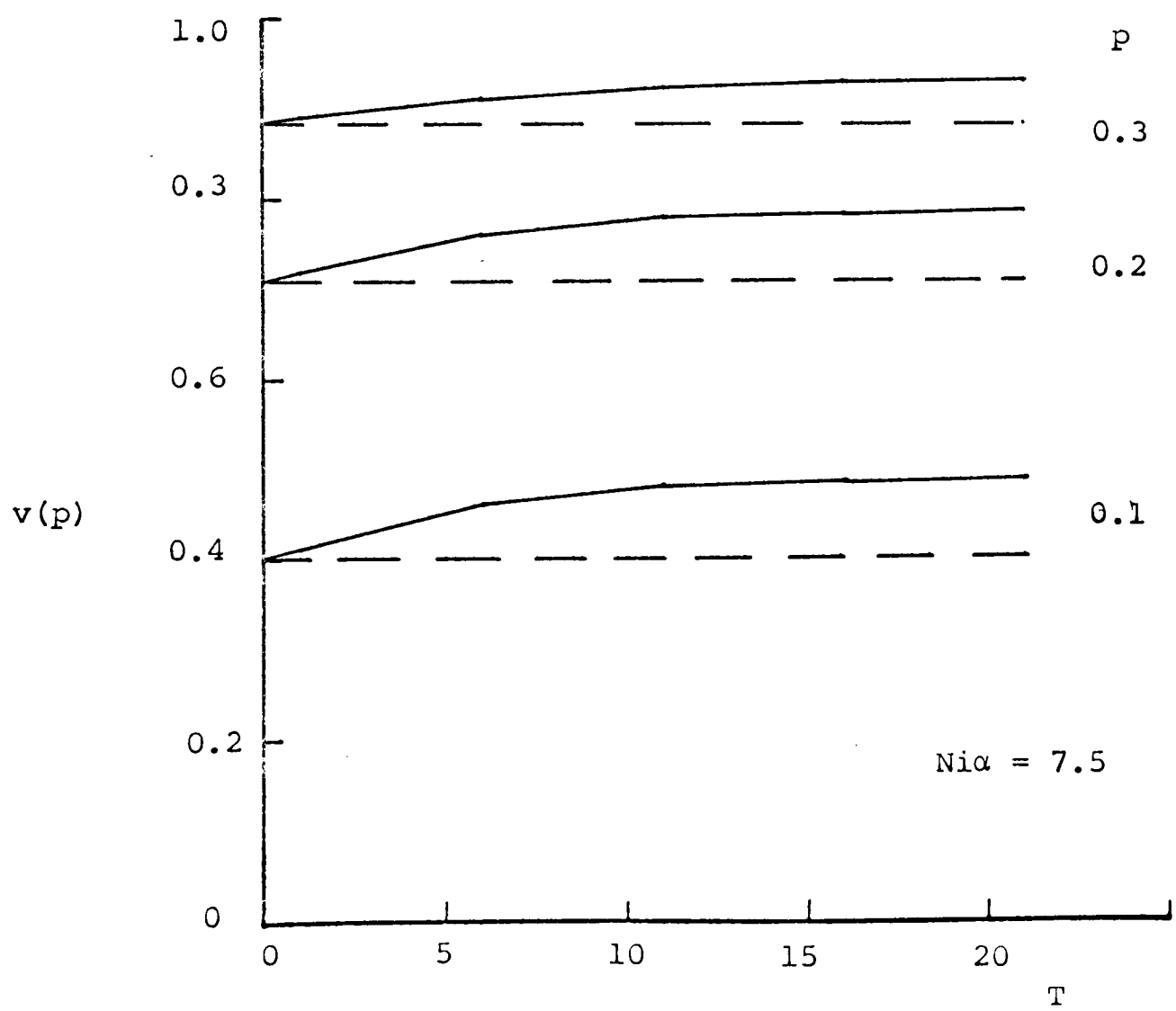


FIGURE 5.4. The relationship between chance of fixation under a sub-lining system and the number of generations of sub-lining for N=10, n=1, n=0.5, y=0.5 for various values of p and Niα under a recessive model.

of unequal size are selected to fixation the large line may have a moderately high chance of being fixed favourably while the small one may have only a low chance. In the cross the favourable allele is therefore most likely to come only from the large line and will have a reasonably high chance of fixation for either y = h or y = 0.5. However if the favourable allele does come from the smaller line only, it will not have much chance of becoming fixed for y = h but may have a reasonable chance if y = 0.5.

ii) for recessive loci, equal sub-lines selected to fixation and crossed with y = 0.5 gives a higher chance of fixation than equivalent single line selection. Gains are of the order of 8% and there are some problems:

(a) Again if sub-lining is for a short period the expected gain is reduced although in this case it is not reversed.

(b) If p is initially high small losses may be incurred.

To see how this comes about consider once more a gene of quite high selective advantage initially at low frequency. Selection in lines of size N/2 tends to give only a small reduction in chance of fixation, since it would be low anyway, but crossing to give a situation where the frequency is 0, 0.5 or 1 gives a relatively large increase.

So although a higher chance of fixation can be obtained in the single locus case the differences observed are small and variable in size and involve a considerable period of extra selection time to achieve. Therefore it is unlikely that any of these rather special sub-lining systems will be of any practical value.

The next problem to be considered concerns linkage, does this make any difference to the conclusions reached above for additive loci? This question will be examined in the next two sections.

(c) Two locus models

The initial base population is assumed to be in linkage equilibrium for two additive loci with alleles A/a and B/b segregating.

Initially compare simple mass selection with splitting the population into two sub-lines each of half the size, selecting to the limit, crossing equally and reselecting to a new limit with the total size always N.

u(p) = chance of fixation of A under the single line system;

u(q) = " " B " " "

v(p) = " " A " sub-lined "

v(q) = " " B " " "

u(p) and u(q) can only in general be obtained by simulation, before considering how v(p) and v(q) are obtained some other quantities must be defined, let,

$u(f_i) = f_i$ = the chance of fixation of the gametes in the single line size N

$u(g_i) = g_i$ = the chance of fixation of the gametes in the sub-line size N/2.

e = chance of fixation of A segregating alone at a frequency of 0.5 in a population size N.

d = chance of fixation of B segregating alone at a frequency of 0.5 in a population size N.

Both these are given by single locus theory.

f_A^+ and f_B^+ = the chance of fixation of A and B respectively when selected in a population of size N initially in extreme positive linkage disequilibrium with frequencies at 0.5.

f_A^- and f_B^- = as above for initial extreme negative linkage disequilibrium

$w(p) = g_1+g_2$ = the chance of fixation of A in the sub-line size N/2

$w(q) = g_1+g_3$ = the chance of fixation of B in the sub-line size N/2.

Then v(p) and v(q) are given by

$$v(p) = (g_1 + g_2)^2 + 2g_1g_4\, f_A^+ + 2g_2g_3\, f_A^- + 2e(g_1g_3 + g_2g_4)$$

$$v(q) = (g_1 + g_3)^2 + 2g_1g_4\, f_B^+ + 2g_2g_3\, f_B^- + 2d(g_1g_2 + g_3g_4)$$

In general the g_i must be obtained by simulation. This is also true for the f_A^+, f_B^+, f_A^- and f_B^- unless Nc = 0 in which case these can be given by single locus theory. It is this extreme case which will be considered first.

1) No Recombination

Nc = 0 then

$$f_A^+ = f_B^+ = \frac{1}{1+e^{-Ni(\alpha+\beta)}}$$

$$f_B^- = 1 - f_B^- = \frac{1}{1+e^{-Ni(\alpha-\beta)}}$$

$$e = \frac{1}{1+e^{-Ni\alpha}}\ ,\ d = \frac{1}{1+e^{-Ni\beta}}$$

$$\therefore\ v(p) = (g_1+g_2)^2 + \frac{2g_1g_4}{1+e^{-Ni(\alpha+\beta)}} + \frac{2g_2g_3}{1+e^{-Ni(\alpha+\beta)}} + \frac{2(g_1g_3+g_2g_4)}{1+e^{-Ni\alpha}} \qquad \ldots(56)$$

In order to see the effect of sub-lining and recrossing simulation was done for the case where Nc = 0 to obtain values of f_i and g_i and so to compare v(p), as given above, with u(p). This was done for a variety of parameter sets, initially for $\alpha=\beta$ and p=q, so that u(p)=u(q) and v(p)=v(q). Figure 5.5 shows chance of fixation plotted against p giving u(p), v(p) and w(p) together with typical ranges plus or minus two standard errors. In no cases are there any large differences between u(p) and v(p) although for some cases where a large number of replicates have been run significant differences have been found. It is difficult to draw any general conclusions from these results but there emerges a tendency for v(p) to exceed u(p) for cases where w(p) is in the region of 0.5. However in general it appears that linkage makes no appreciable difference to the conclusion reached for independent loci, namely that this form of sub-lining makes no difference to the ultimate chance of fixation.

The apparent superiority of the sub-lining system when w(p) is in the region of 0.5 bears further investigation. For $\alpha=\beta$ and p=q, v(p) is given by

$$v(p) = (g_1+g_2)^2 + \frac{2g_1g_4}{1+e^{-2Ni\alpha}} + g_2g_3 + \frac{2(g_1g_3 + g_2g_4)}{1+e^{-Ni\alpha}}$$

$$= w(p) + \frac{1-e^{-2Ni\alpha}}{1+e^{-2Ni\alpha}}\ g_1g_4 + \frac{1-e^{-Ni\alpha}}{1+e^{-Ni\alpha}}\ (g_1g_3 + g_2g_4) \qquad \ldots (57)$$

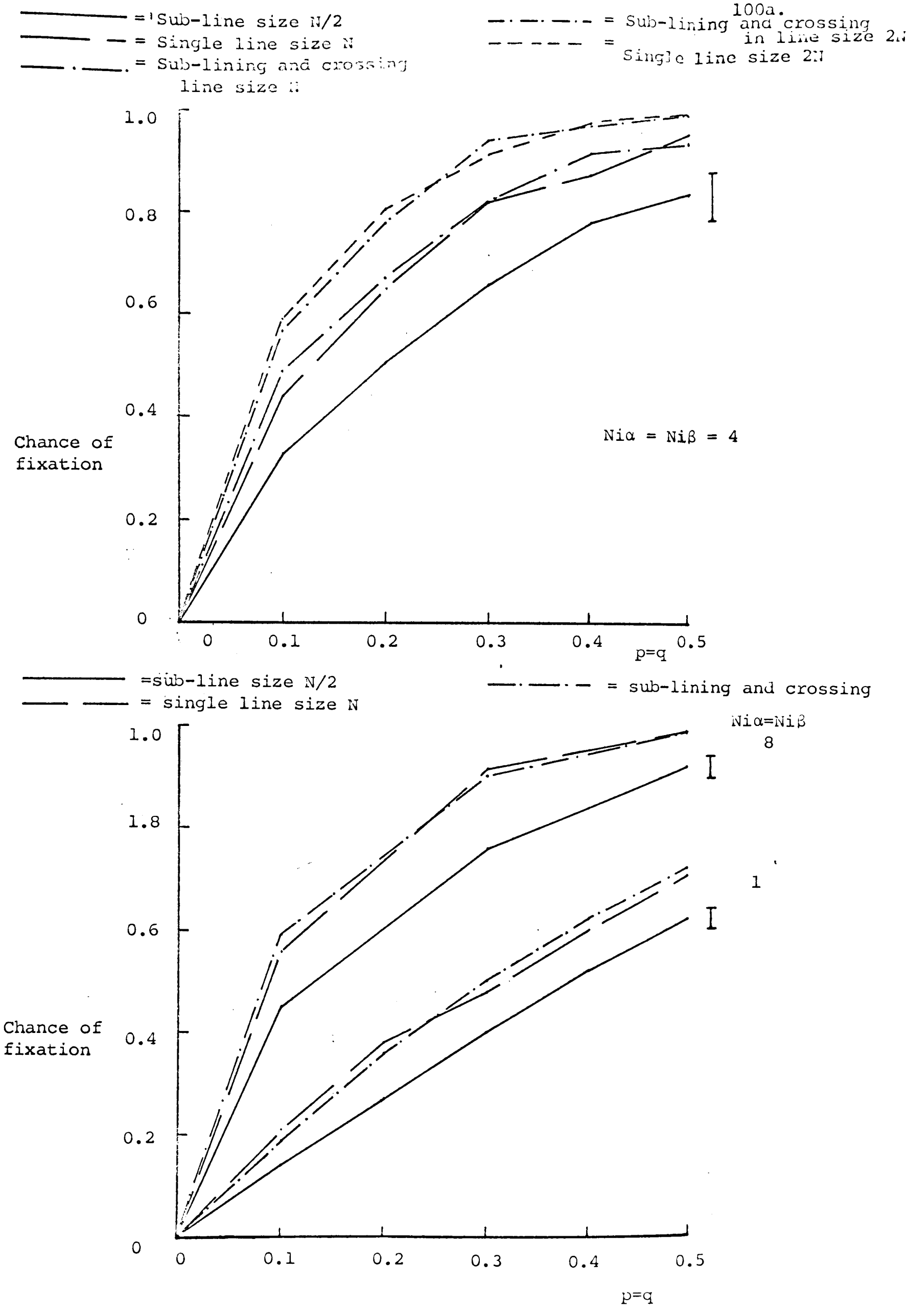


FIGURE 5.5. The effect of sub-lining and crossing for N=8, n=2, Niα = Niβ, Nc=0, h=0.5, y=0.5, p=q, with selection from a single base population. Typical ranges of length four standard errors are also shown.

$v(p) - w(p)$ gives the increase in chance of fixation due to crossing the fixed lines.

This, after some rearrangement, can be given as

$$v(p) - w(p) = \frac{1-e^{-Ni\alpha}}{1+e^{-Ni\alpha}} \left[w(p)\ (1-w(p)\ ) + \frac{2e^{-Ni\alpha} D_B}{1+e^{-2Ni\alpha}} - \frac{(1-e^{-Ni\alpha})^2}{1+e^{-2Ni\alpha}} (w(p)(1-w(p)\ ) - D_B)^2 \right] \quad \ldots (58)$$

where $D_B = g_1 g_4 - g_2 g_3$ = the disequilibrium between lines at the limit.

If $Ni\alpha$ is large such that $e^{-Ni\alpha} \simeq 0$ this approximates to

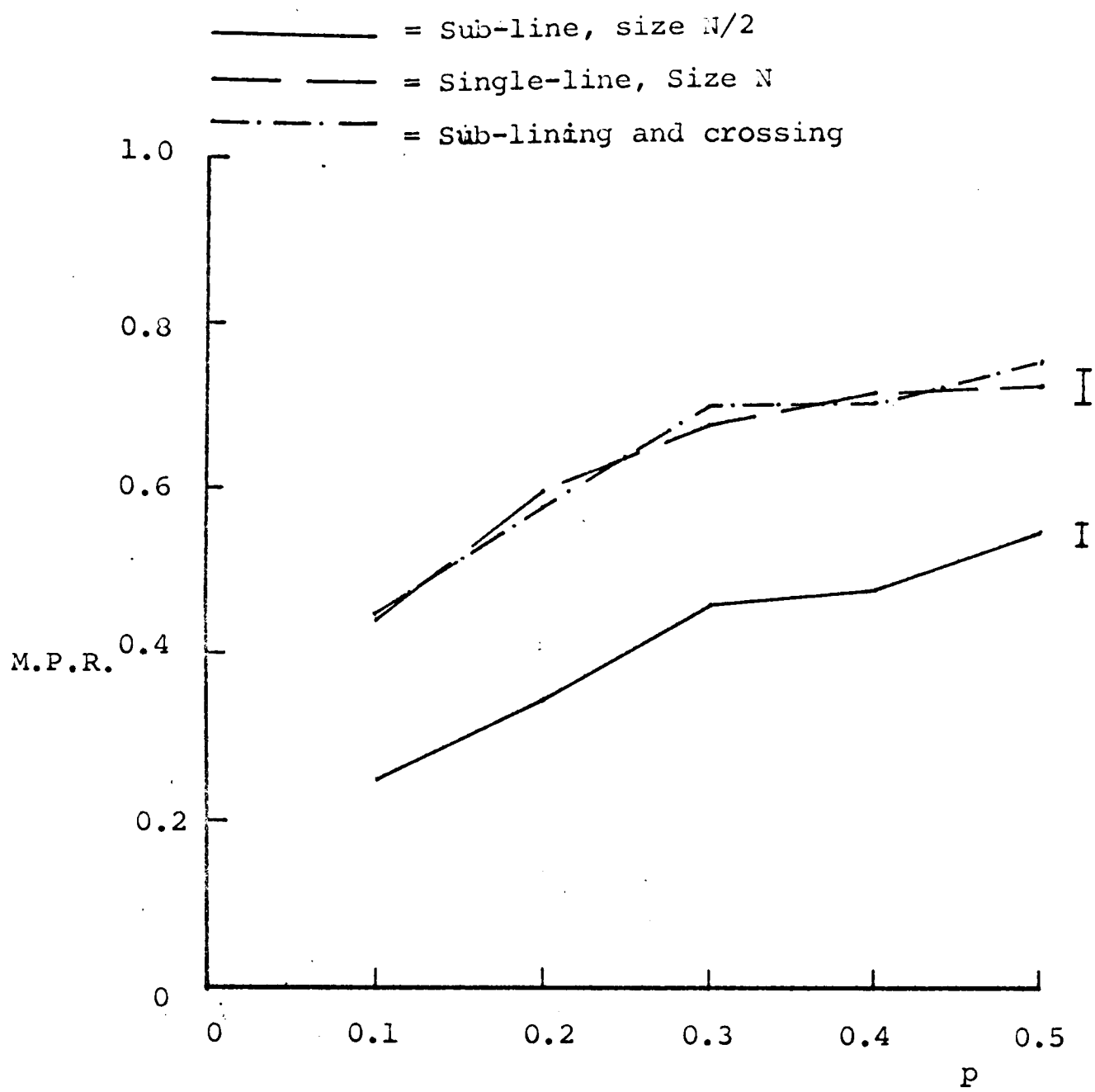
$$v(p) - w(p) = w(p)(1-w(p)) - (w(p)(1-w(p)) - D_B)^2 \quad \ldots (59)$$

Selection for linked loci has been shown to generate negative linkage disequilibrium between lines (Hill and Robertson 1966) so that an increase in the magnitude of the disequilibrium gives a decrease in the gain expected due to crossing. Examination of a wide variety of simulation results obtained under initial linkage equilibrium has revealed that D_B is negative and large, of the order of -0.1, for loci of large effect with chance of fixation around 0.5 - 0.7. However for loci of small effect D_B is generally small although still negative, if $D_B \simeq 0$ then $v(p) - w(p)$ is a function solely of $Ni\alpha$ and $w(p)(1-w(p))$, being maximized for $w(p) = 0.5$. For large loci $-D_B$ also increases as $w(p)$ approaches 0.5 and this may affect the maximum to a small extent.

Next consider the effect of sub-dividing the population in the first place, this is given by u(p) - w(p) but it cannot be formulated for the linked locus case. However simulation results indicate that although this is in general of very similar magnitude to v(p) - w(p) it is not maximized for w(p) = 0.5 such that at this point v(p) > u(p). Even for the simple case of complete linkage with equal effects and initial frequencies it has not been possible to adequately explain the observed results, except to say that v(p) - w(p) appears to be rather more dependent on the product w(p)(1-w(p)) than does u(p) - w(p) but differences are nevertheless always small. To see how inequality of gene effects and frequencies affects this situation more simulation results will be presented. Figure 5.6 shows results for N = 8, Niα = 4, Niβ = 2, q = 0.3, Nc = 0, with the chance of fixation and M.P.R. plotted against p. This does not give a very clear picture although for all p there are no significant differences between u(p) and v(p), while differences do appear between u(q) and v(q) with in general v(q) > u(q). This is probably because A being the larger effect locus is little affected by the linked B and so behaves as if independent, this is equivalent to putting $e^{-Ni(\alpha+\beta)} = e^{-Ni(\alpha-\beta)} = e^{-Ni\alpha}$ in equation (56) then

$$v(p) - w(p) \approx \frac{1-e^{-Ni\alpha}}{1+e^{-Ni\alpha}} \; w(p)\,(1-w(p))$$

which is the same as for independent loci. However using the same approximations in v(q) gives



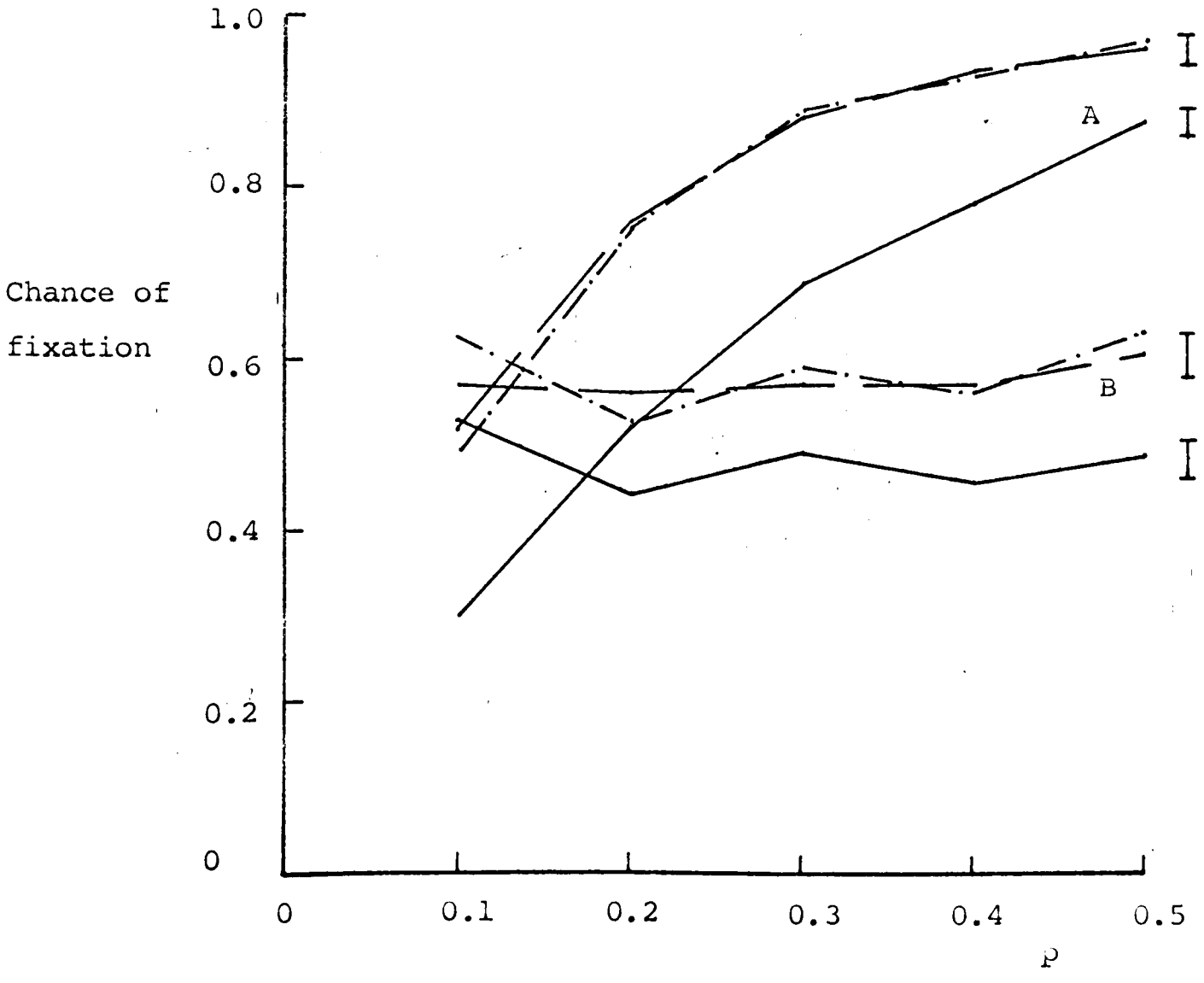


FIGURE 5.6. The effect of sub-lining and crossing for N=8, n=2, Niα =4, Niβ=2, Nc=0, h=0.5, y=0.5, q=0.3 for various values of p, with selection from a single base population. Typical ranges of length four standard errors are also shown.

$$v(q) - w(q) = \frac{1-e^{-Ni\beta}}{1+e^{-Ni\alpha}} (w(q)(1-w(q)) - [w(p)-(1-w(q))-D_B]$$
$$[w(q)(1-w(p))-D_B] + \frac{1-e^{-Ni\alpha}}{1+e^{-Ni\alpha}} D_B \quad \ldots (60)$$

ie. linkage has a very marked effect on the smaller locus, considerably reducing $u(q) - w(q)$ particularly for $w(p) = 0.5$. However linkage obviously has a similar effect on $u(q) - w(q)$ and when the M.P.R. is examined no real differences are found between the sublining and single line systems.

Further results are given in Figure 5.7 for N=8, Nc=0, p=0.2, q=0.3, Niβ=2 with chance of fixation and M.P.R. plotted against iα, differences are again mainly non-significant.

These results suggest that for completely linked loci initially in linkage equilibrium there is little difference between sub-lining and crossing versus single line selection. However $v(p)$ and $u(p)$ are not in general identical. The gain in chance of fixation due to crossing given by $v(p) - w(p)$ tends to be influenced mainly by the product $w(p)(1-w(p))$ but is also dependent upon the disequilibrium built up between lines.

ii) Recombination

For a value of Nc between zero and infinity the f_A^+, f_B^+, f_A^- and f_B^- can only be obtained by simulation so that further chance of fluctuations are introduced in the calculations of $v(p)$.

Recombination makes the values of both f_A^+ and f_A^- approach that of e and so causes $v(p) - w(p)$ to approach $\frac{1-e^{-Ni\alpha}}{1-e^{-Ni\alpha}} w(p)(1-w(p))$ thereby tending to increase it from the Nc = 0 case. Recombination also increases $w(p)$ for the same

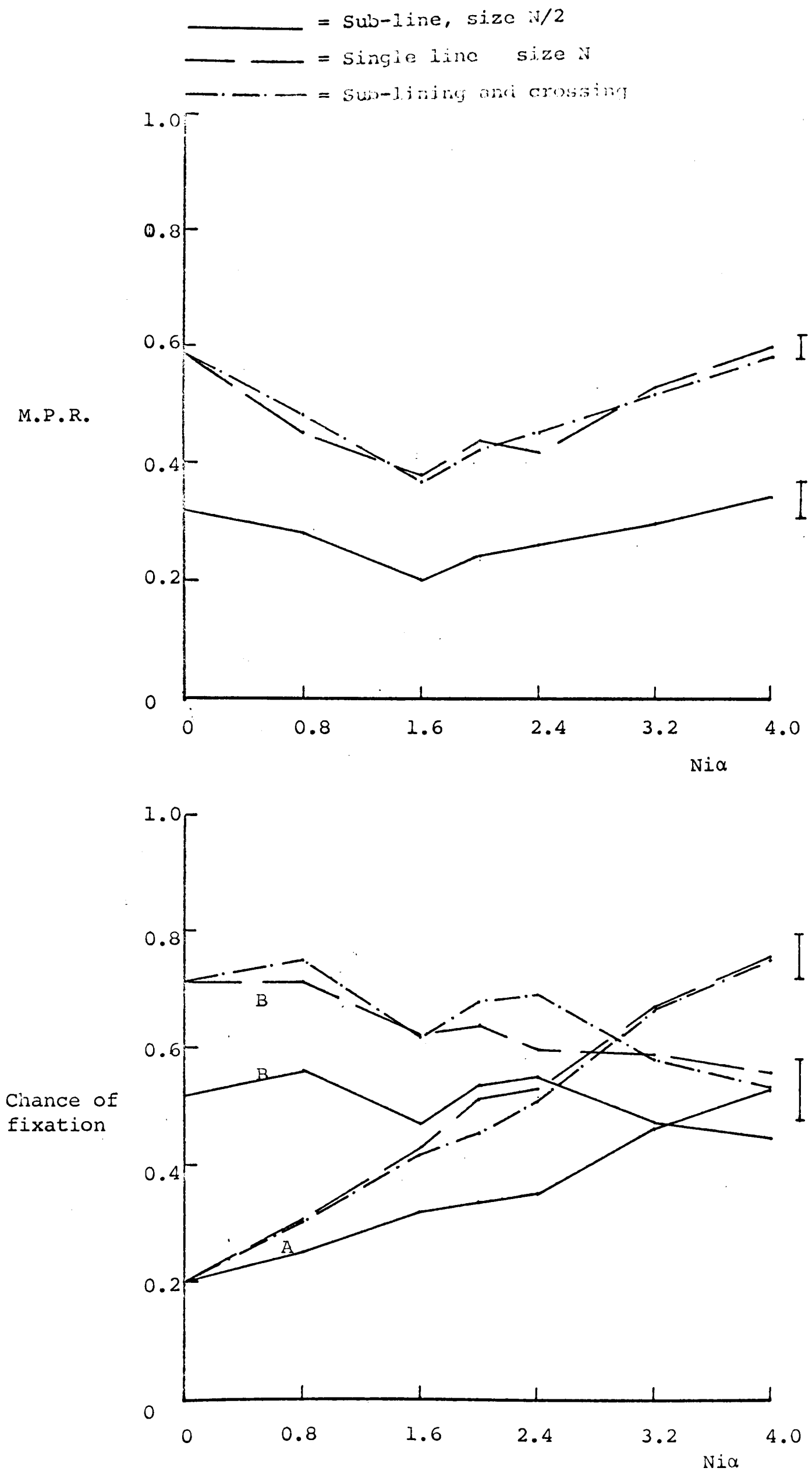


FIGURE 5.7. The influence of inequality of gene effects on the difference between single line and sub-line systems of selection for N=8. n=2, Niß=2, Nc=0, p=0.2, q=0.3, h=0.5, y=0.5 with selection from a single base population. Typical ranges of length four standard errors are also shown.

effects and initial frequencies. Simulation results are shown in Figure 5.8 for the equal effects and initial frequencies case, with chance of fixation plotted against $\frac{Nc}{2(1+2Nc)}$. For all values of Nc, w(p) ≈ 0.5 and there is a tendency for v(p) > u(p) but all differences are small and non-significant.

Therefore recombination does not alter the observation that in general there is no useful difference between sub-lining and crossing and single line selection, for a pair of additive loci.

Consider the effect of altering the system slightly in order to make it more realistic, instead of selecting the sub-lines to fixation suppose they were selected for only T generations, and then crossed and reselected. In this case the only comparison is v(p) with u(p) and since simulation alone can be used the two locus model has no advantages, therefore this has been studied under the multi locus model.

(d) Multi locus models

In this case a four locus model has been simulated with equal gene frequencies, q, and equal recombination fractions c. Initially the aim has been to see if sub-lining for some T generations followed by crossing has any effect on the limit. Figure 5.9 shows results of simulation for N=10, Niα at all loci = 5, Nc=0, q=0.2, with the mean frequency plotted against generation number. This shows that there are no significant differences at the limit for T=0,5, 15 and 25, it also shows that there are no significant differences in initial rate of response,

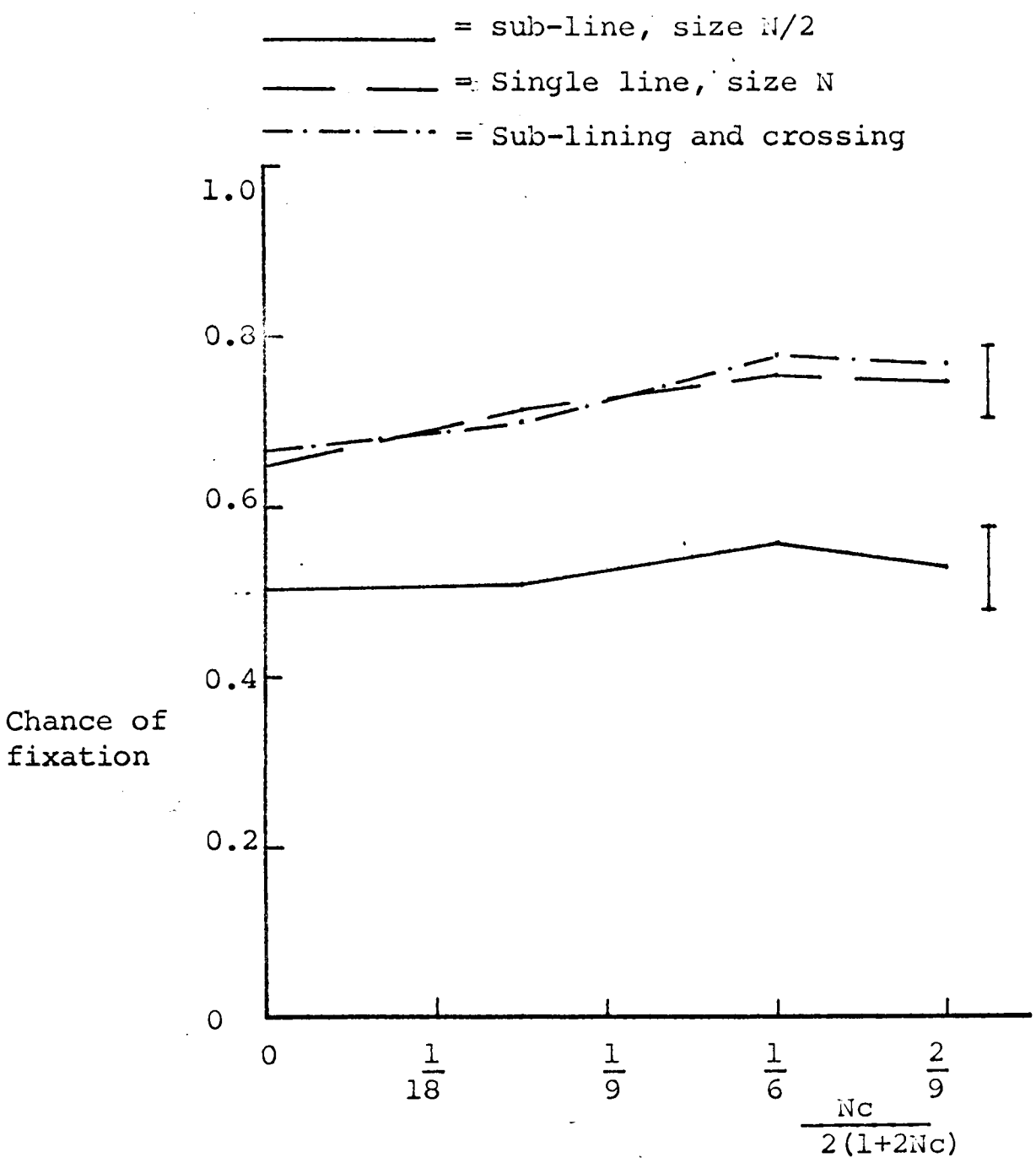


FIGURE 5.8. The influence of recombination on the effect of sub-lining and crossing for N=8, n=2, Niα=Niβ=4, p=q=0.2, h=0.5, y=0.5 with selection from a single base population. Typical ranges of length four standard errors are also shown.

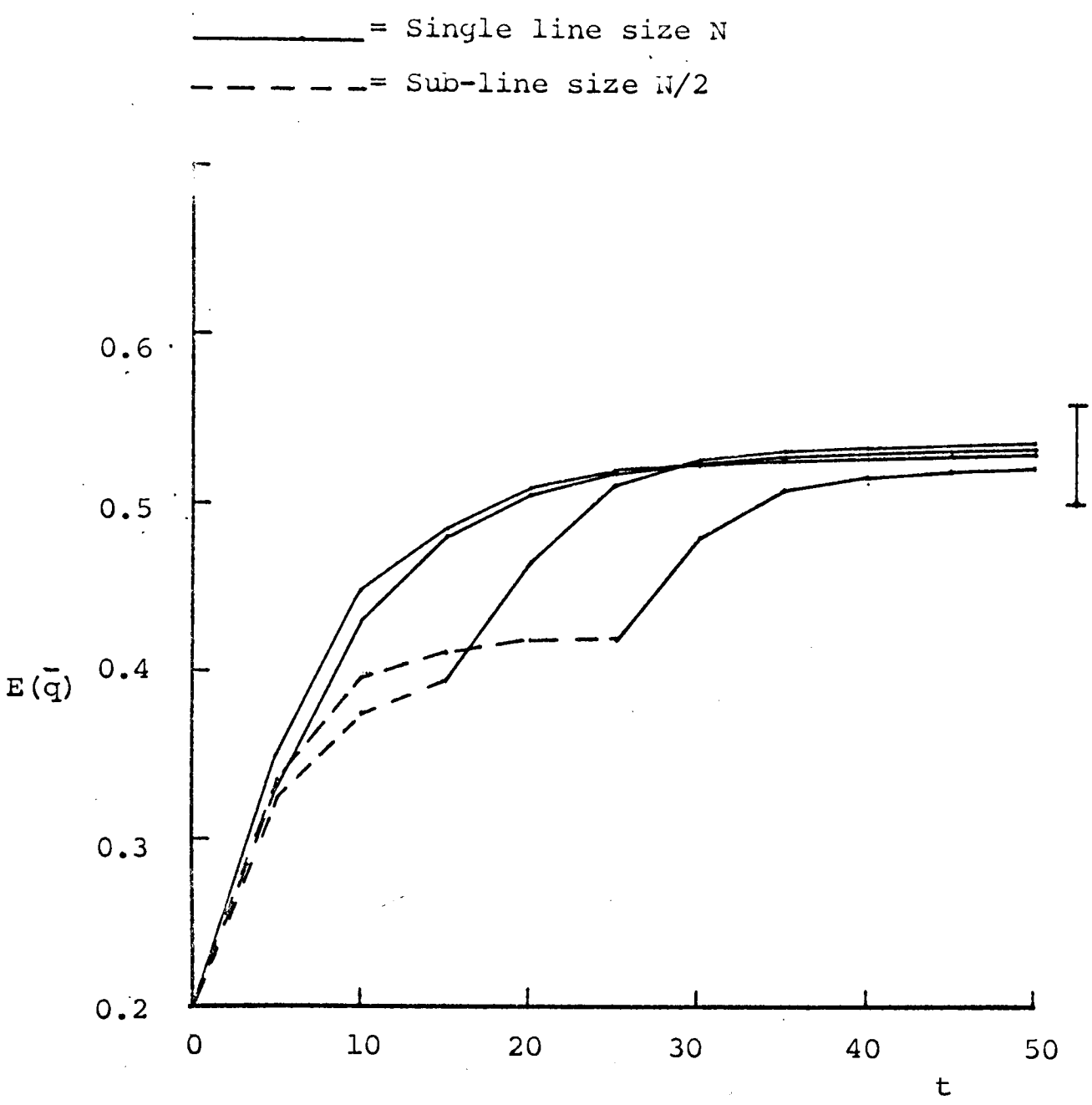


FIGURE 5.9. The rate of response under single line and sub-line systems of selection for N=8, n=4, Niα=5, Nc=0, $\bar{q}$=0.2, h=0.5, y=0.5 with selection from a single base population. Typical range of length four standard errors is also shown.

i.e. over the first five generations. However the sub-lining systems show a markedly reduced mean frequency after say 20 generations and it is this fact which makes sub-lining impractical in general.

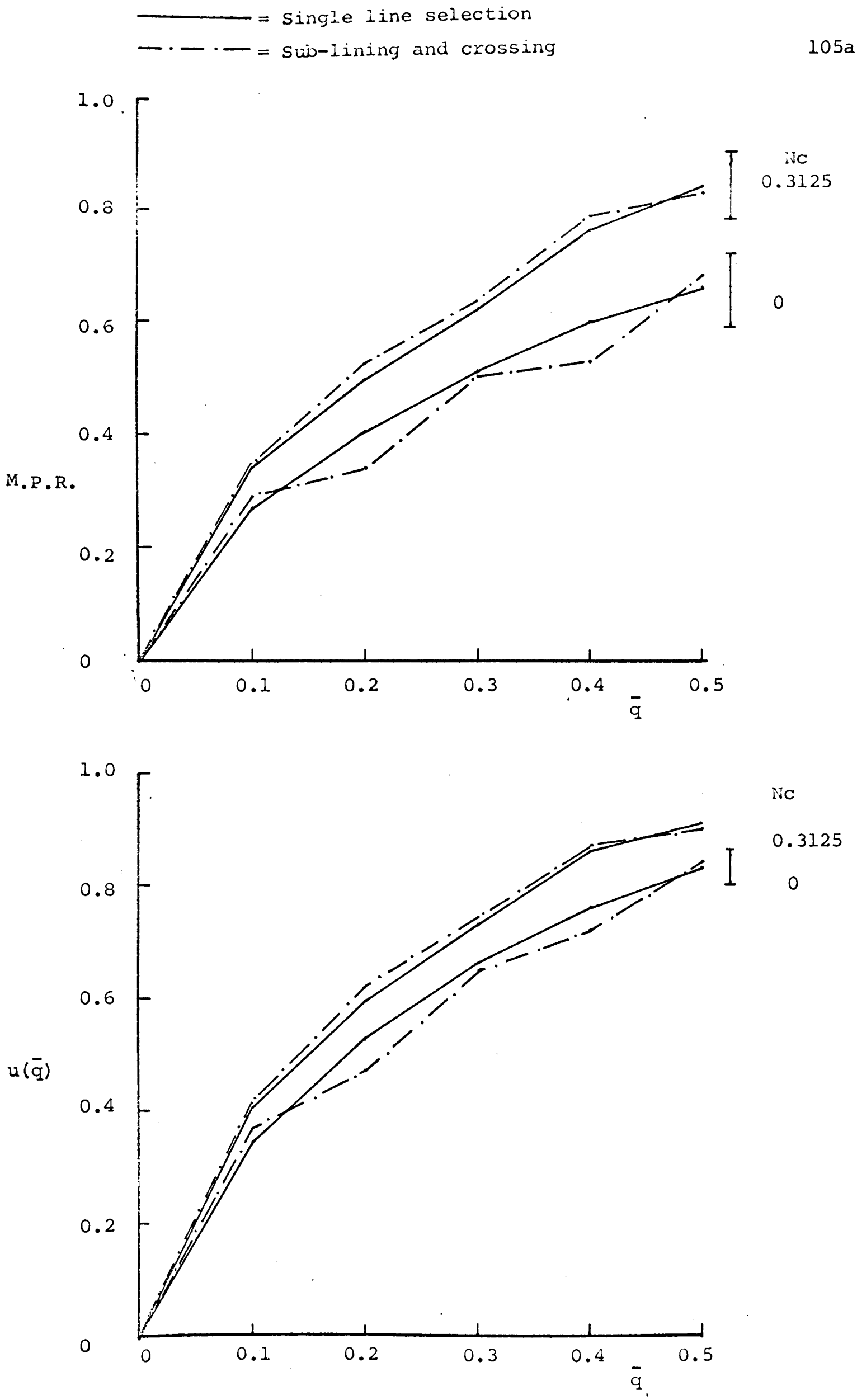
Further simulation results are presented in Figures 5.10 and 5.11 for N=10, T=10, Nc=0 and 0.3125, $N_i\alpha$ = 2.5, 5 and 7.5, q=0.1, 0.2, 0.3, 0.4 and 0.5. These show that most differences are again non-significant and suggest that in no cases will there be any useful advantage in the sub-lining system.

Some simulation was also done for one of the cases considered under the single locus model, that of sub-dividing the population into unequal sub-lines, selecting for T generations and crossing with y=0.5. Results for h=0.2 are shown in Figure 5.12 for T=10 and 20. They suggest that for T=10 sub-lining is on the whole worse than the single line system but for T=20 there is no apparent difference, which is what might have been expected on the basis of the two locus results.

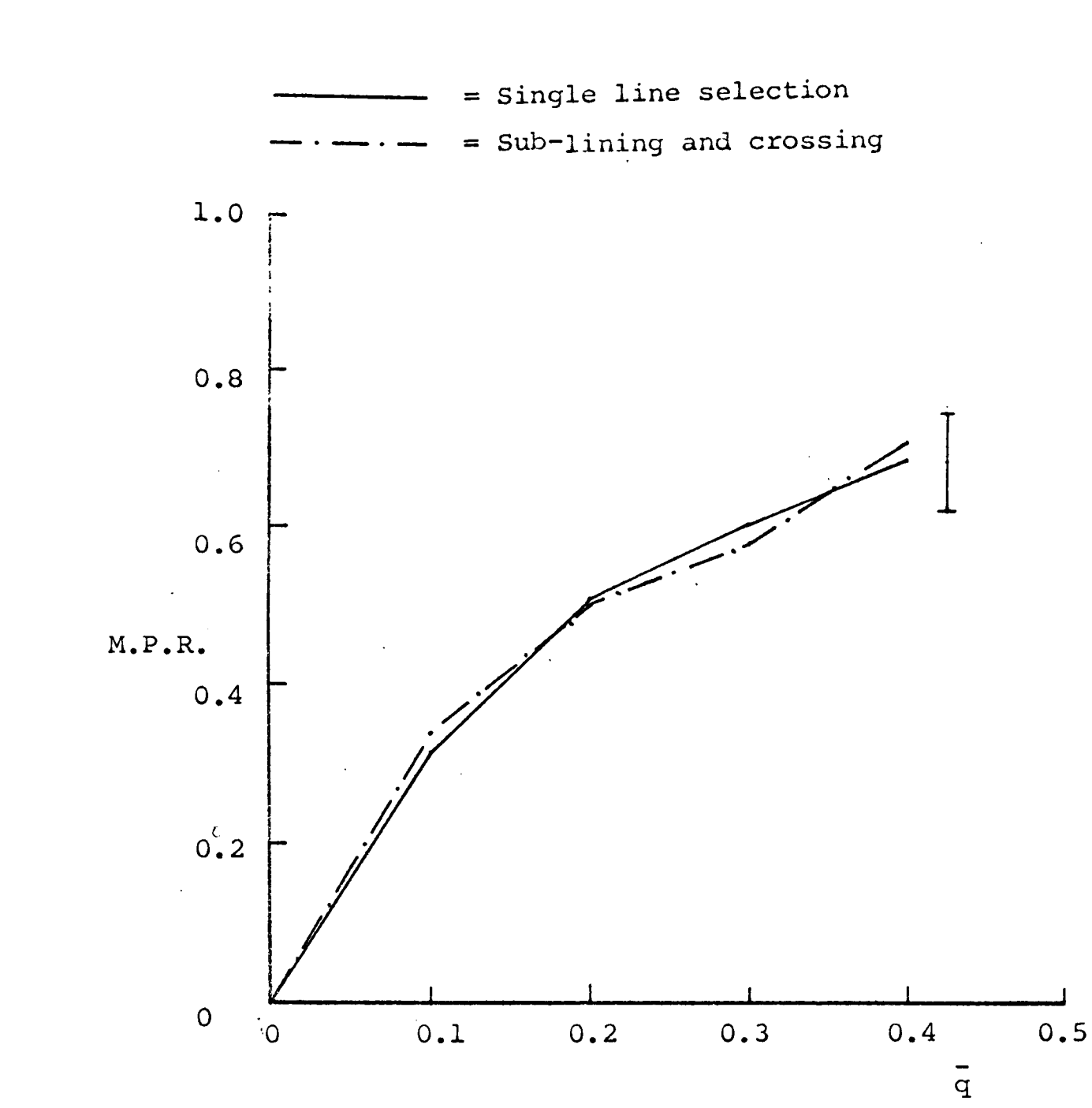
(e) Discussion and Conclusions.

Although some of the studies under the very restricted single and two locus models suggest that sub-lining and crossing may increase chance of fixation and therefore ultimate response it has been found that in general these differences are very small.

Madalena (1970) has studied essentially the same problem under a multilocus model. He compared rates of response and chance of fixation under a single line system with various sub-line selection strategies. He found that if there was no selection between



IGURE 5.10. The effect of sub-lining and crossing under a multilocus model for N=10, n=4, Niα=5, h=0.5, y=0.5, T=10 with selection from a single base population. Typical ranges of length four standard errors are also shown.



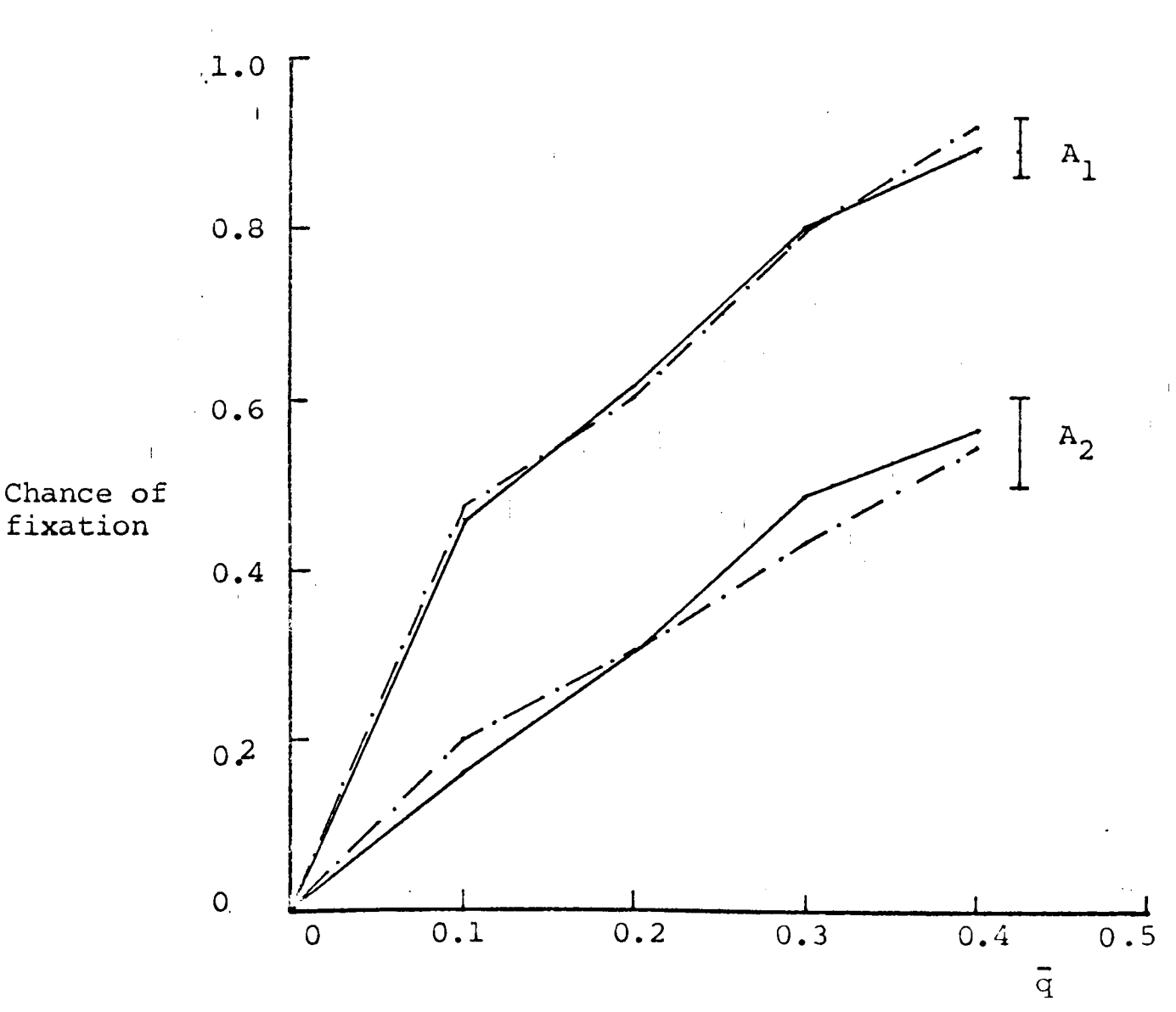


FIGURE 5.11. The effect of sub-lining and crossing under a multi-locus model with unequal effects for N=10, n=4, Niα=7.5, Niα_2=2.5, h=0.5, y=0.5 T=10 with selection from a single base population. Typical ranges of length four standard errors are also shown.

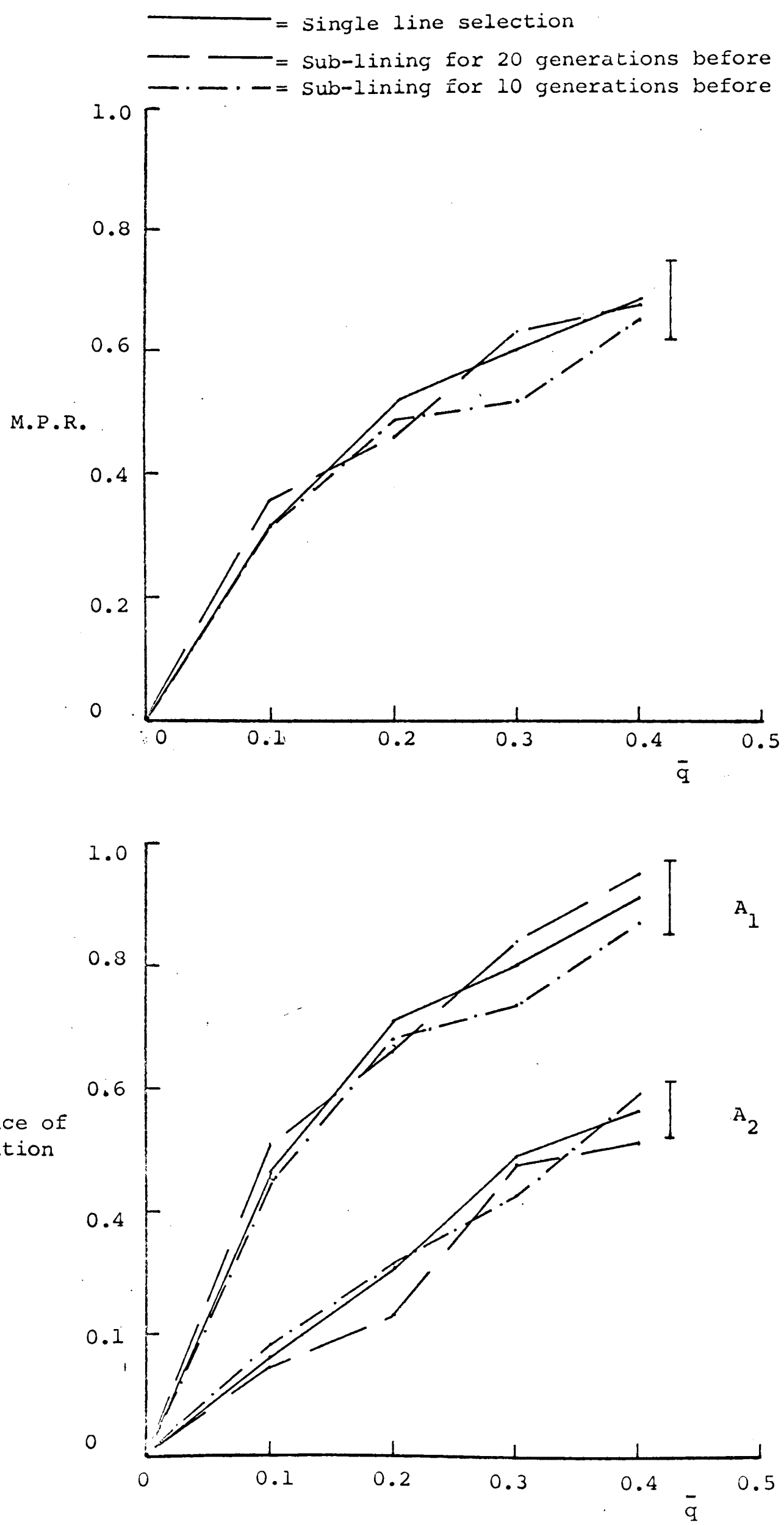


FIGURE 5.12. The effect of sub-lining and crossing under a multi-locus model with unequal effects for N=10, n=4, $Ni\alpha_1$=7.5, $Ni\alpha_2$=2.5, Nc=0 h=0.2, y=0.5 for T=10 and 20 with selection from a single base population. Typical ranges of length four standard errors are also shown.

sub-lines then the limit was not affected by the sub-lining although the rate of response was considerably reduced.

The results presented here for the case involving linked loci in general confirm the findings of Madalena with respect to chance of fixation. However the two locus studies do suggest that there may in fact be some small differences between single line and sub-lining systems and that the magnitude and sign of these may be dependent upon the chance of fixation of the genes in the sub-line.

Nevertheless under the multi locus model there has been no sign of the sub-lining system being consistently better in terms of chance of fixation, under any circumstances. However restrictions on computing time has limited the number of replications made under the multi-locus model to the region of 100 and this does not give sufficient accuracy to detect differences of the magnitude predicted under the simpler models.

Since sub-lining has a detrimental effect on rate of response, particularly for intermediate generations, it seems unlikely that any of the sub-lining systems discussed here will be of any practical value in increasing overall response. However, this study has also shown that if sub-lining is desirable for some entirely independent reason, the limit reached by a comparable single line can still be attained simply by crossing the selected sub-lines.

CHAPTER VI

Selection Procedures with distinct base Populations

(a) Introduction

A question which may of considerable importance in practical breeding concerns the best way to utilize variation from several distinct populations. Previous work on this topic has been reviewed in Chapter Two. This suggests that if several base populations are available to the breeder then selection within rather than between populations may be desirable. In view of this conclusion a study has been made to consider some different ways in which separate populations might be combined and selected to reach a maximum limit.

The model studied assumes that there are two distinct base populations available whose genetic means can be accurately determined, and that in all cases a total of N individuals is selected from M. In particular the following questions have been considered:

i) Should only one population be used, if so on what basis should such a choice be made?

ii) If both populations are to be used to form a base population, in what proportions should each contribute?

iii) If a cross population is to be made with say hN individuals from one population and (1-h)N from the other, should these be selected separately as sub-lines first, or should they be crossed at the outset?

As in the previous chapter three basic models have been considered although the single locus model is better regarded as an independent locus model, since although what is generally termed 'single locus theory' is used it is applied to two loci assumed to be segregating entirely independently.

(b) Independent loci

Suppose that the only difference between the two populations is in the frequency of a single locus, then the population with the higher mean will be the population with the gene at the higher frequency. If the means are equal then this gives the single population case, otherwise the best that can be done is to select entirely from the population with the higher mean, this applies under additive or dominance models.

The situation becomes more complex if two independent loci are considered. Let the two loci be A and B and let

the frequency of A in population 1 = p_1

" B " 1 = q_1

" A " 2 = p_2

" B " 2 = q_2

If A and B have effects α and β respectively, then the mean in population 1, $M1 = p_1 \alpha + q_1 \beta + K$ and the mean in population 2, $M2 = p_2 \alpha + q_2 \beta + K$ where K is due to other loci for which both populations are assumed to be genetically identical.

Questions (i) and (ii) can be considered together by letting hN = the number of individuals taken from population 1

(1-h)N= " " " " " 2

where h = 0, $\frac{1}{N}$ ... $\frac{N-1}{N}$, 1, and ask what value of h will maximize the mean at the limit. However first question (iii) will be considered since this can be answered, for the additive case, for any value of h. Let

$\bar{p}$ = the frequency of A in the cross = $hp_1 + (1-h)p_2$

$\bar{q}$ = " " B " " = $hq_1 + (1-h)q_2$

$u(\bar{p})$ = the chance of fixation of A if selection is in the cross made at the outset

$u(\bar{q})$ = as above for B.

then

$$u(\bar{p}) = \frac{1-e^{-2Ni\alpha(hp_1 + (1-h)p_2)}}{1-e^{-2Ni\alpha}}$$

and

$$u(\bar{q}) = \frac{1-e^{-2Ni\beta(hq_1 + (1-h)q_2)}}{1-e^{-2Ni\beta}}$$

Consider the effect of selecting in sub-lines to fixation before crossing, then

$u(p_1)$ = chance of fixation of A selected in the sub-line taken from population 1

$u(q_1)$ = as above for B

$u(p_2)$ = chance of fixation of A selected in the sub-line taken from population 2

$u(q_2)$ = as above for B

$$u(p_1) = \frac{1-e^{-2Ni\alpha hp}}{1-e^{-2Ni\alpha h}} \quad , \quad u(p_2) = \frac{1-e^{-2Ni\alpha(1-h)p_2}}{1-e^{-2Ni\alpha(1-h)}}$$

Let

$v(p)$ = the chance of fixation of A under the sub-lining and crossing system

$v(q)$ = as above for B.

Then $v(p) = u(p_1) \times u(p_1) \times 1 + u(p_2) \times (1-u(p_2)) \times \frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}}$

$$+ u(p_2)(1-u(p_1)) \times \frac{1-e^{-2Ni\alpha(1-h)}}{1-e^{-2Ni\alpha}}$$

$$= \frac{1-e^{-2Ni\alpha(hp_1 + (1-h)p_2)}}{1-e^{-2Ni\alpha}} \qquad = u(\bar{p})$$

That is selection in sub-lines followed by crossing makes no difference to the limit. This result is again predicted by that of Maruyama (1970).

This simple relation no longer holds for dominance models since chance of fixation can only be evaluated numerically. This has been done for $h = 0.1, 0.2 \ldots 0.9$ and p_1 and $p_2 = 0.1, 0.2 \ldots 0.9$ and u and v compared. In all cases differences were small, being greatest for h = 0.5 then

$v(p) > u(\bar{p})$ for a recessive if $\bar{p} \leqslant 0.6$ approx

$v(p) < u(\bar{p})$ " " $\bar{p} > 0.6$ "

$v(p) < u(\bar{p})$ for a dominant if $\bar{p} \leqslant 0.6$ "

$v(p) > u(\bar{p})$ " " $\bar{p} > 0.6$ "

which is similar to the results found for the single locus case. For cases where $h \neq 0.5$ it was found that the main factor in determining the sign of $v(p) - u(\bar{p})$ was the frequency in the population which contributed least to the cross. This result is generally reasonable since changes in frequency are likely to have most effect on chance of fixation in the smaller sub-line. For example for h = 0.1 it was found that

$v(p) > u(\bar{p})$ for a recessive if $p_1 \leqslant 0.6$ approx.

$v(p) < u(\bar{p})$ for a dominant if $p_1 \leqslant 0.6$ approx.

In all cases differences were very small but in general

$v(p) > u(\bar{p})$ for a recessive and

$v(p) < u(\bar{p})$ for a dominant.

Next questions (i) and (ii) will be considered in terms of the value of h which will maximize the mean at the limit. Let ML = the contribution of these loci to the mean at the limit, for the additive case

$$ML = u(\bar{p})\ \alpha + u(\bar{q})\ \beta$$

the derivative of ML with respect to h is given by

$$\frac{d}{dh} ML = 2Ni\left[\frac{\alpha^2(p_1-p_2)e^{-2Ni\alpha(hp_1 + (1-h)p_2)}}{1-e^{-2Ni\alpha}} + \frac{\beta^2(q_1-q_2)e^{-2Ni\beta(hq_1 + (1-h)q_2)}}{1-e^{-2Ni\beta}}\right]$$

if $p_1 = p_2$ and $q_1 = q_2$, $\frac{d}{dh}ML = 0$ for all h.

i.e. if the populations are genetically identical it makes no difference in what proportions they are crossed. If $p_1 > p_2$ and $q_1 > q_2$ $\frac{d}{dh} ML$ is always positive and so the maximum value of ML which can be achieved is given by h = 1, i.e. selecting only from population 1. Similarly if $p_2 > p_1$ and $q_2 > q_1$, ML is maximized for h = 0, i.e. selecting only from population 2. Otherwise the value of h which will maximize ML is given by

$$h_{max} = \frac{\left[\log_e \left[\frac{\beta^2}{\alpha^2} \left(\frac{q_2 - q_1}{p_1 - p_2}\right)\left(\frac{1-e^{-2Ni\alpha}}{1-e^{-2Ni\beta}}\right)\right] - 2Ni\ (\beta q_2 - p_2\alpha)\right]}{2Ni\ (p_2\ \alpha - p_1\ \alpha + q_1\ \beta - q_2\beta)} \quad \ldots(61)$$

but this does not necessarily give $0 < h < 1$ so for $h_{max} < 0$, $h = 0$ must be used and for $h_{max} > 1$, $h = 1$ must be used. For $\alpha = \beta$ the above simplifies to

$$h_{max} = \frac{\log_e \left(\frac{q_2 - q_1}{p_1 - p_2}\right) - 2Ni\alpha\ (q_2 - p_2)}{2Ni\alpha\ (p_2 - p_1 + q_1 - q_2)} \quad \ldots\ (62)$$

In the previous chapter it was shown that for the sub-lining and crossing system chance of fixation could in some cases be increased by crossing in a proportion different from that in which the original sub-lining was done. In this case consider having sub-lines of size hN and (1-h)N, with h as above, but crossing after selection in the proportions y:(1-y). Let chance of fixation from this system be $v_y(p)$ and $v_y(q)$, with $D_y(p) = v_y(p) - v(p)$ and $D_y(q) = v_y(q) - v(q)$. Then

$$v_y(p) = u(p_1)u(p_2) + u(p_1)(1-u(p_2))\ \frac{1-e^{-2Ni\alpha y}}{1-e^{-2Ni\alpha}}$$

$$+ u(p_2)(1-u(p_1))\ \frac{1-e^{-2Ni\alpha(1-y)}}{1-e^{-2Ni\alpha}}$$

and

$$D_y(p) = u(p_1)(1-u(p_2))\ \left(\frac{e^{-2Ni\alpha}\ \underline{h}\ e^{-2Ni\alpha y}}{1-e^{-2Ni\alpha}}\right)$$

$$+ u(p_2)(1-u(p_1))\ \left(\frac{e^{-2Ni\alpha(1-h)} - e^{-2Ni\alpha(1-y)}}{1-e^{-2Ni\alpha}}\right)$$

$$\ldots\ (63)$$

$D_y(p) + D_y(q)$ gives the difference in ML due to the disproportionate crossing. If $y = h\,D_y(p) + D_y(q) = 0$ otherwise this is maximized for y_{max} given by

$$y_{max} = \frac{\log_e \left[\frac{u(p_1)(1-u(p_2)) + u(q_1)(1-u(q_2))}{u(p_2)(1-u(p_1)) + u(q_2)(1-u(q_1))} \right]}{4\,Ni\alpha} + \frac{1}{2} \qquad \ldots (64)$$

which approaches $\frac{1}{2}$ for large $Ni\alpha$.

In practice α, β, p_1, p_2, q_1 and q_2 will not be known so that these equations for h_{max} and y_{max} may not be very useful. All that will be known are the population means M1 and M2, so consider two cases: (i) M1 = M2 and (ii) M1 > M2.

(i) M1 = M2

If it is assumed that no more than this is known then one of the following courses of action might reasonably be taken:

(a) Select entirely from population 1, i.e. put h = 1

(b) Select entirely from population 2, i.e. put h = 0

(c) Cross the populations equally, i.e. put h = 0.5

Since there is no reason to choose population 1 more often than population 2, a comparison of chance of fixation can be made between:

I. Selecting from a single population chosen at random

II. Selecting from the cross with h = 0.5

Let

$\overline{u(p)}$ = the chance of fixation of A with selection from a single population chosen at random

$$\therefore \quad \overline{u(p)} = \left(\frac{1-e^{-2Ni\alpha p_1}}{1-e^{-2Ni\alpha}} + \frac{1-e^{-2Ni\alpha p_2}}{1-e^{-2Ni\alpha}} \right) / 2$$

$$= \frac{2-e^{-2Ni\alpha p_1} - e^{-2Ni\alpha p_2}}{2(1-e^{-2Ni\alpha})} \qquad \ldots(65)$$

$$u(\bar{p}) - \overline{u(p)} = \frac{e^{-2Ni\alpha p_1} - 2e^{-Ni\alpha(p_1 + p_2)} + e^{-2Ni\alpha p_2}}{2(1-e^{-2Ni\alpha})}$$

$$= \frac{(e^{-2Ni\alpha p_1} - e^{-2Ni\alpha p_2})^2}{2(1-e^{-2Ni\alpha})} \qquad \ldots(66)$$

which is positive for all p_1 and p_2 if $Ni\alpha > 0$, unless $p_1 = p_2$ in which case the populations are genetically identical and $u(\bar{p}) = \overline{u(p)}$.

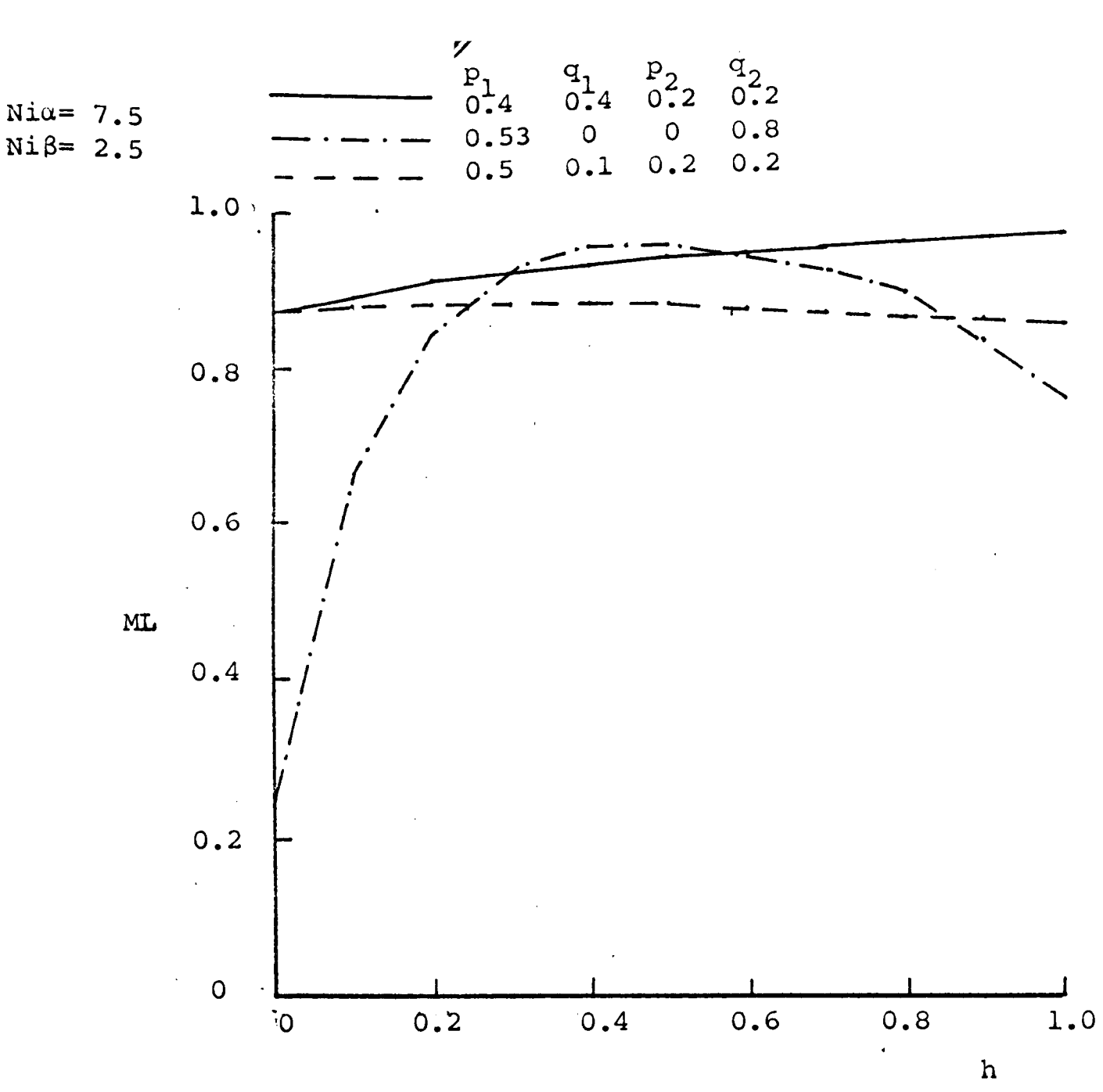
Therefore in general the highest chance of fixation for both loci is given by making the 50:50 cross.

ii) M1 > M2

The way in which ML varies with h is in general complex and depends on α and β and the relative values of p_1, p_2, q_1 and q_2.

Examples for M1 = 0.4 + K, M2 = 0.2 + K and N=10 are shown in Figure 6.1. These show that under some situations h = 1 gives the highest M2 while under others a quite low value of h gives a clear advantage.

A special case worth considering further is that when the population with the greater mean is fixed, say by previous selection, while the population with the lower mean is still segregating at both loci,; for example let $\alpha = \beta$, $p_2 = q_2 = p$ and $p_1 = 0$, $q_1 = 1$ then M1 = α + K, M2 = 2α p + K



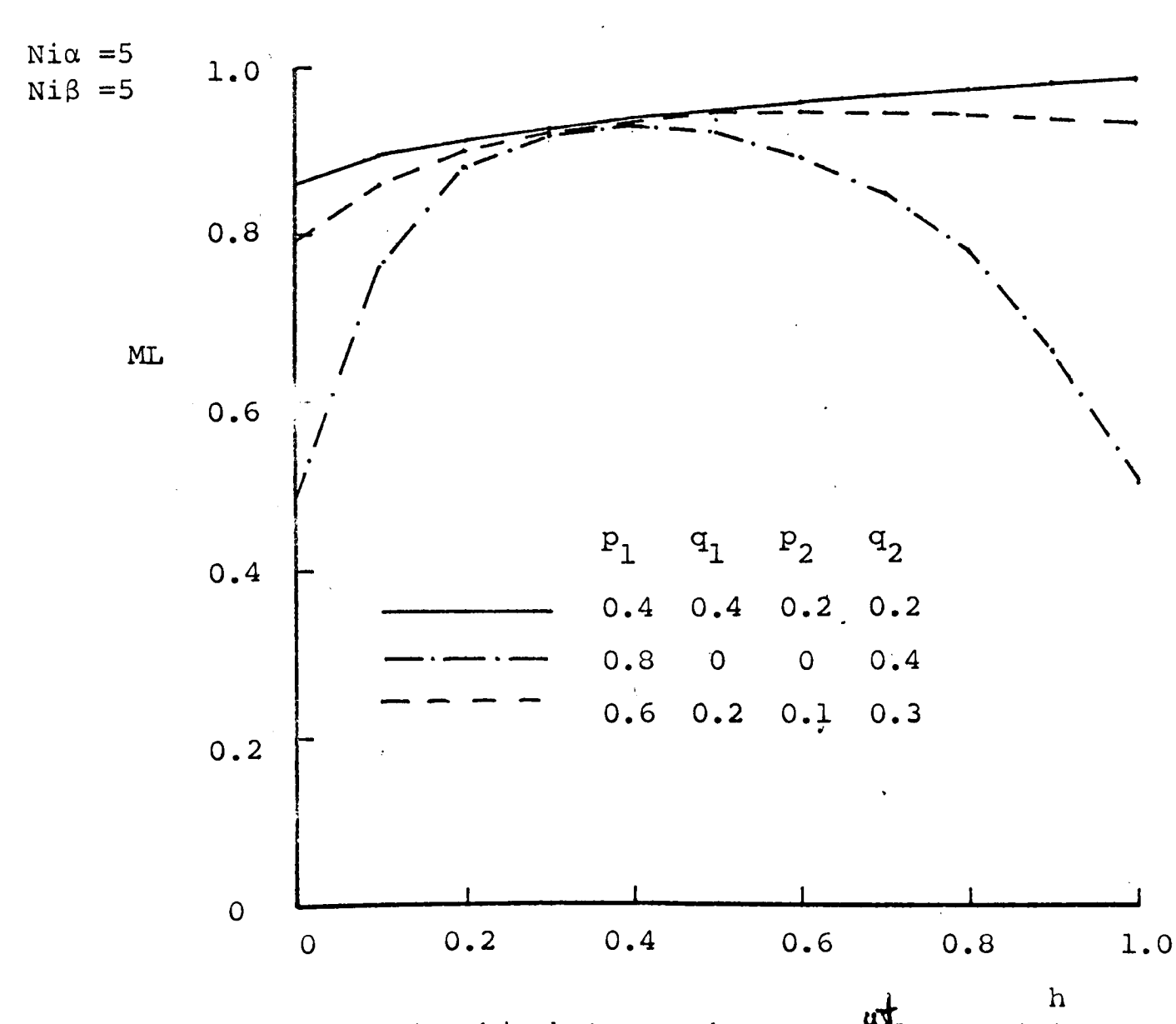


FIGURE 6.1. The relationship between the mean and the limit and h for independent additive loci for N=10, n=2 under various models of initial frequency distributions, such that M1=0.4 and M2=0.2.

$\therefore$ M1 - M2 = α (1-2p), so if M1 > M2, p < 0.5

If a cross is made at the outset, with hN individuals from the fixed population then

$$u(\bar{p}) = \frac{1-e^{-2Ni\alpha(1-h)p}}{1-e^{-2Ni\alpha}} \quad , \quad u(\bar{q}) = \frac{1-e^{-2Ni\alpha(h+(1-h)q)}}{1-e^{-2Ni\alpha}}$$

$$\text{and } ML = \alpha \frac{[2 - e^{-2Ni\,(1-h)p} - e^{-2Ni\,(h+(1-h)p)}]}{1-e^{-2Ni\alpha}} \qquad \ldots (67)$$

this is maximized for $h = \frac{\log_e(\frac{1}{p} - 1)}{2Ni\alpha}$

Therefore the larger the effects at the loci concerned the greater the proportion which should be taken from the segregating population. Also the closer p is to 0.5 the nearer h becomes to zero, so that if the difference between the means is small it will be in general better to select entirely from the segregating population. However if differences are large due to loci expected to be at low frequency it might be better not to make the cross at all. It is in this type of situation which it might be useful to use a system involving sub-line selection with proportions h and (1-h) followed by crossing with proportions y and (1-y). Since the first population is fixed there is clearly no advantage in selecting a sub-line from this population, therefore put h = 0, then $u(p_1) = 0$, $u(p) = \frac{1-e^{-2Ni\alpha p}}{1-e^{-2Ni\alpha}}$ and $u(q_1) = 1$.

$$\therefore y_{max} = \frac{\log_e[\frac{1}{u(p)} - 1]}{4Ni\alpha} + \frac{1}{2}$$

which again approaches 0.5 if Niα is large and u(p) intermediate.

This type of selection system has been discussed by Osman and Robertson (1968). The chance of fixation for the two loci is given in general by

$$v_y(q) = u(p) \times 1 + (1-u(p)) \times \frac{1-e^{-2Ni\alpha y}}{1-e^{-2Ni\alpha}}$$

$$v_y(p) = u(p) \left(\frac{1-e^{-2Ni\alpha(1-y)}}{1-e^{-2Ni\alpha}}\right)$$

$$ML = \alpha[v_y(p) + v_y(q)]$$

$$= \alpha\left[\frac{u(p)\ (1-e^{-2Ni\alpha(1-y)})(1+e^{-2Ni\alpha y}) + 1-e^{-2Ni\alpha y}}{(1-e^{-2Ni\alpha})}\right] \quad \ldots(68)$$

if we cross equally after selection, i.e. put $y = 0.5$ this gives

$$ML = \alpha\left[\frac{2-e^{-2Ni\alpha p} - e^{-Ni\alpha}}{1-e^{-2Ni\alpha}}\right] \quad \ldots(69)$$

In general this system will give a higher mean at the limit unless the difference between the populations is due to small genes at very low frequency. Then for the previous system h=1 would have been used to give $M1 = \alpha$, i.e. the mean of the fixed population, in this case even selection of the segregating population before crossing may serve to reduce M1.

Another special case of interest is where both populations are fixed but for different loci such that $M1 = \alpha+K$, $M2 = \beta+K$, $p_1=q_2=0$, $p_2=q_1=1$

then

$$u(\bar{p}) = \frac{1-e^{-2Ni\alpha(1-h)}}{1-e^{-2Ni\alpha}} \qquad u(\bar{q}) = \frac{1-e^{-2Ni\beta h}}{1-e^{-2Ni\beta}}$$

$$ML = \alpha \left[\frac{1-e^{-2Ni\alpha(1-h)}}{1-e^{-2Ni\alpha}} + \frac{\beta}{\alpha} \frac{1-e^{-2Ni\beta h}}{1-e^{-2Ni\beta}} \right] \qquad \ldots(70)$$

this is maximized for

$$h = \frac{\log_e [\frac{\beta^2}{\alpha^2} (\frac{1-e^{-2Ni\alpha}}{1-e^{-2Ni\beta}})] + 2Ni\alpha}{2Ni(\alpha+\beta)} \qquad \ldots(71)$$

for $\alpha = \beta$ $h = 0.5$ maximizes ML

if $\alpha > \beta$ then ML is maximized for some value of h greater than $\frac{\alpha}{\alpha+\beta}$

As $Ni\alpha$ becomes large so this value of h approaches $\frac{\alpha}{\alpha+\beta}$. This can be given in terms of the means M1 and M2 as

$$\frac{\alpha}{\alpha+\beta} = \frac{1}{2 - \frac{(M1-M2)}{\alpha}} .$$

Figure 6.2 shows some examples of the way h_{max} varies with β/α for different values of $Ni\alpha$. If $Ni\alpha > 4$ $h = \frac{\alpha}{\alpha+\beta}$ gives a good approximation to h_{max}. Otherwise using $\frac{\alpha}{\alpha+\beta} \leqslant h \leqslant 0.5$ gives ML in no cases far from the maximum as the second part of Figure 6.2 shows.

Similar comparisons can be made for the non-additive case, again for M1 = M2 and M1 > M2.

i) M1 = M2

Compare again

$$\overline{u(p)} = \frac{u(p_1) + u(p_2)}{2}$$

with

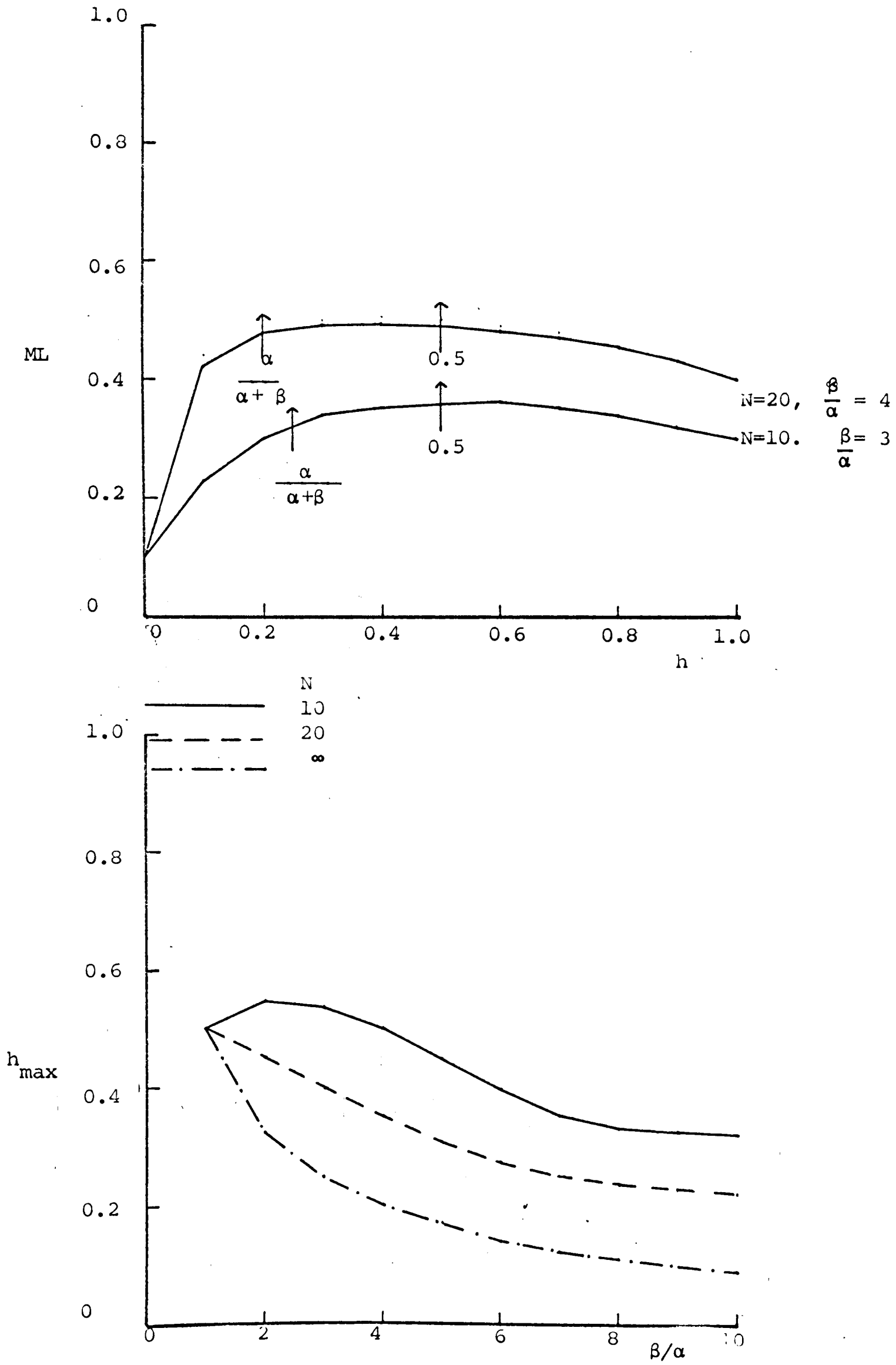


FIGURE 6.2 The relationships between the mean at the limit. and β/ for independent additive loci when the difference in population means is due to differences in effect of the genes carried.

$$u(\bar{p}) = u\left(\frac{p_1 + p_2}{2}\right)$$

(a) selection for a complete recessive

$$\overline{u(p)} = \frac{\int_0^{p_1} e^{-2Ni\alpha x^2}\,dx + \int_0^{p_2} e^{-2Ni\alpha x^2}\,dx}{2\int_0^{1} e^{-2Ni\alpha x^2}\,dx} \qquad \ldots(72)$$

$$u(\bar{p}) = \frac{2\int_0^{\frac{p_1+p_2}{2}} e^{-2Ni\alpha x^2}\,dx}{2\int_0^{1} e^{-2Ni\alpha x^2}\,dx}$$

if $p_1 = p_2$ $\overline{u(p)} = u(\bar{p})$, otherwise let $p_1 > p_2$

then

$u(\bar{p}) > \overline{u(p)}$ if

$$\int_{p_1}^{\frac{p_1+p_2}{2}} e^{-2Ni\alpha x^2}\,dx > \int_{\frac{p_1+p_2}{2}}^{p_2} e^{-2Ni\alpha x^2}\,dx$$

this will be true if $\frac{d}{dx} e^{-2Ni\alpha x^2}$ is negative in the range $p_1 < x < p_2$

$\frac{d}{dx} e^{-2Ni\alpha x^2} = -4Ni\alpha x \; e^{-2Ni\alpha x^2}$ which is negative for all $x > 0$ if $Ni\alpha > 0$.

Therefore for the recessive case $u(\bar{p}) > \overline{u(p)}$.

(b) Selection for a complete dominant

$$\overline{u(p)} = \frac{\int_0^{p_1} e^{-2Ni\alpha x(2-x)}\,dx + \int_0^{p_2} e^{-2Ni\alpha x(2-x)}\,dx}{2\int_0^{1} e^{-2Ni\alpha x(2-x)}\,dx} \qquad \ldots(73)$$

$$u(\bar{p}) = \frac{2\int_0^{\frac{p_1+p_2}{2}} e^{-2Ni\alpha x(2-x)}\,dx}{2\int_0^{1} e^{-2Ni\alpha x(2-x)}\,dx}$$

if $p_1=p_2$, $\overline{u(p)} = u(\bar{p})$, otherwise let $p_1 < p_2$ and it can again be shown that for $Ni\alpha > 0$ $u(\bar{p}) > \overline{u(p)}$

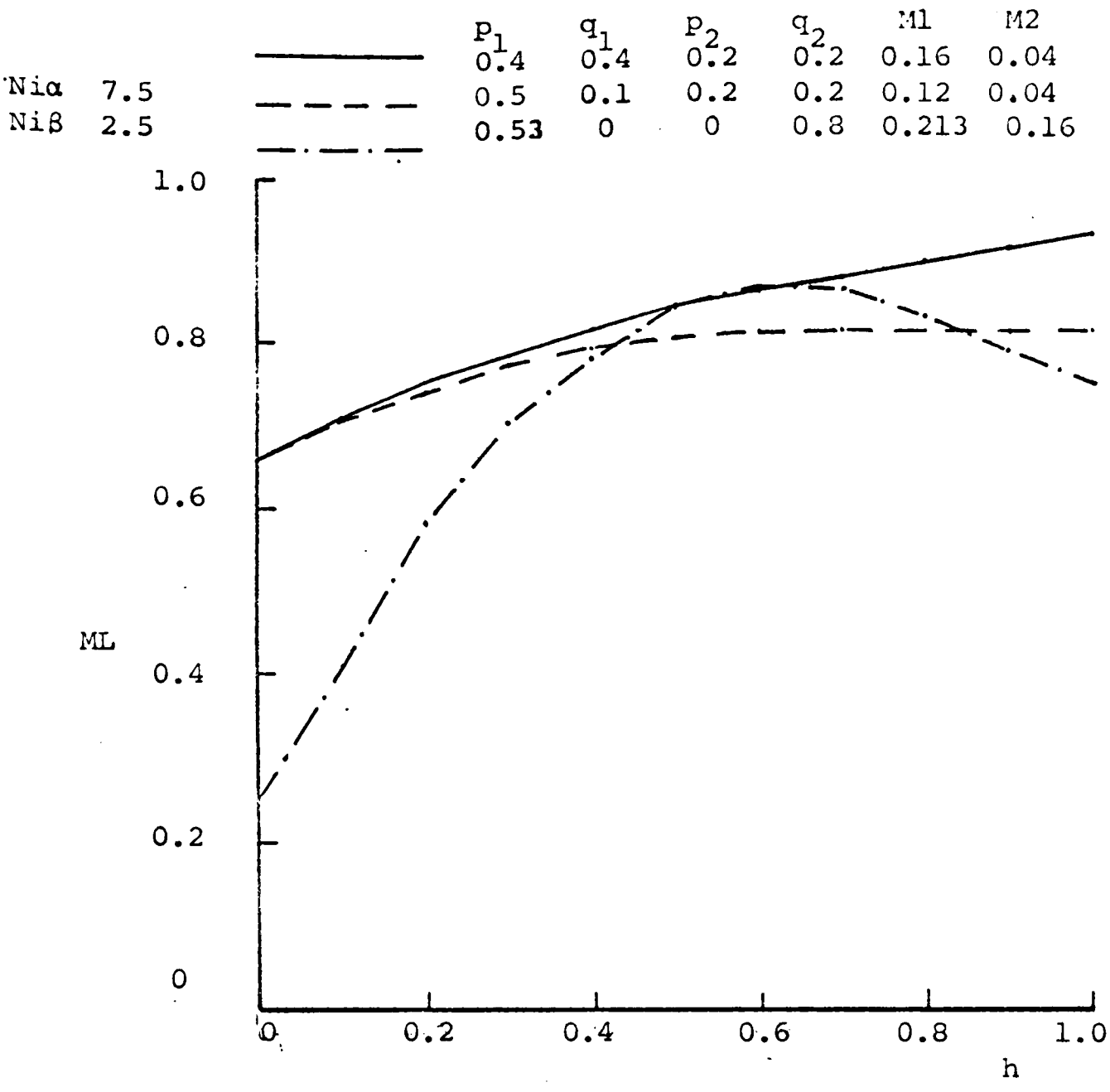
ii) M1> M2

The relationship between ML and h is complex as in the additive case and examples are given in Figures 6.3 and 6.4 for completely recessive and completely dominant loci respectively. The same values of α, β, p_1, q_1, p_2 and q_2 are used as in Figure 6.1 but since

mean = $p(2-p)\alpha + q(2-q)\beta$ for the dominance case and

mean = $p^2\alpha + q^2\beta$ for the recessive case, values of M1 and M2 are not the same although in all cases M1 > M2.

Comparison of figures 6.1, 6.3 and 6.4 reveals that with selection for a recessive the value of ML is very sensitive to changes in h while for an additive it is less so and for a dominant a very wide range of values of h give roughly the same value of ML. This observation is simply a reflection of the way in which chance of fixation responds to changes in population size under the different models of gene action, recessives being most sensitive and dominants much less so. Therefore choosing the right h value may be very important if selection is for recessive loci, conversely if selection is for dominant loci any intermediate value of h may suffice.



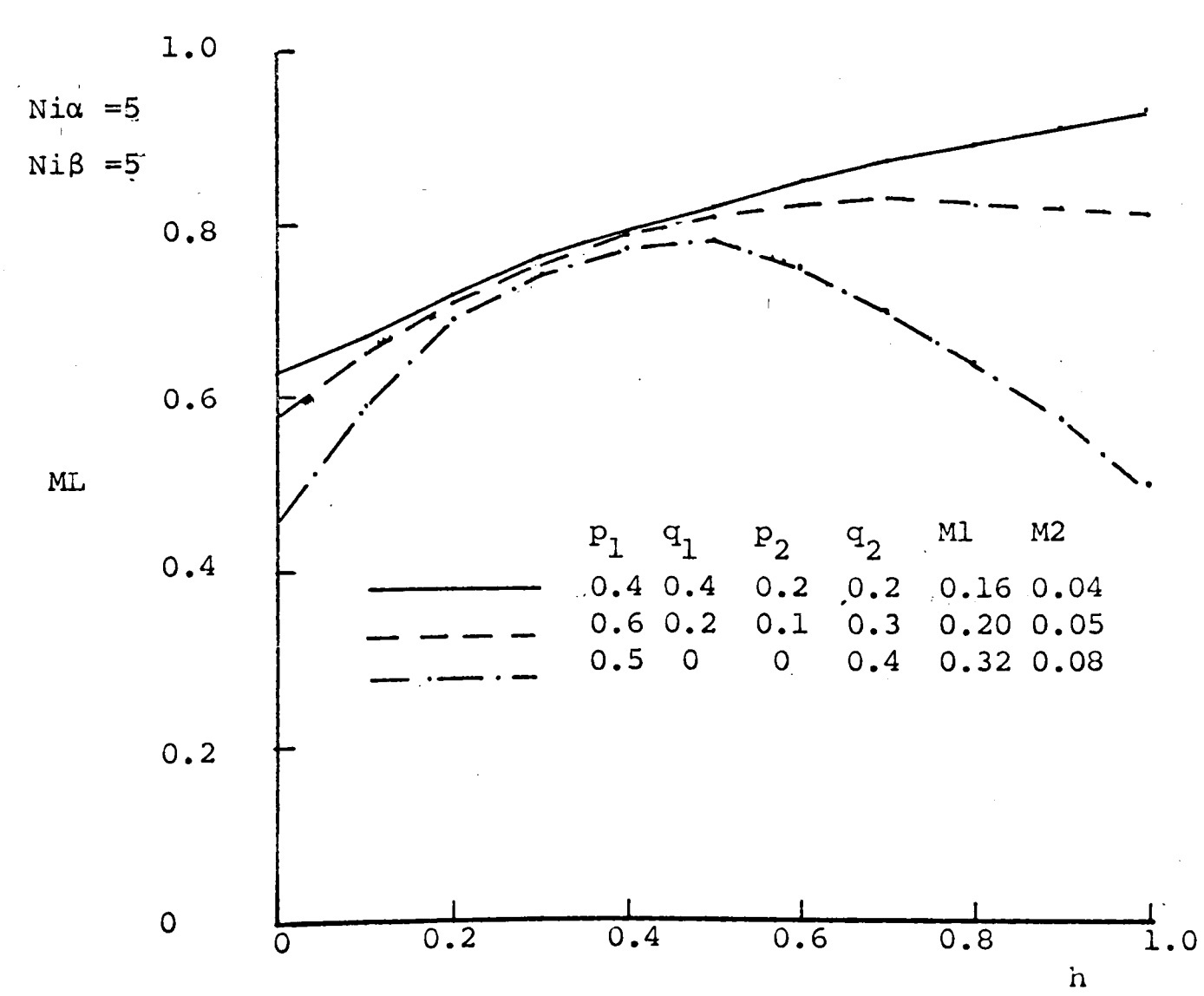
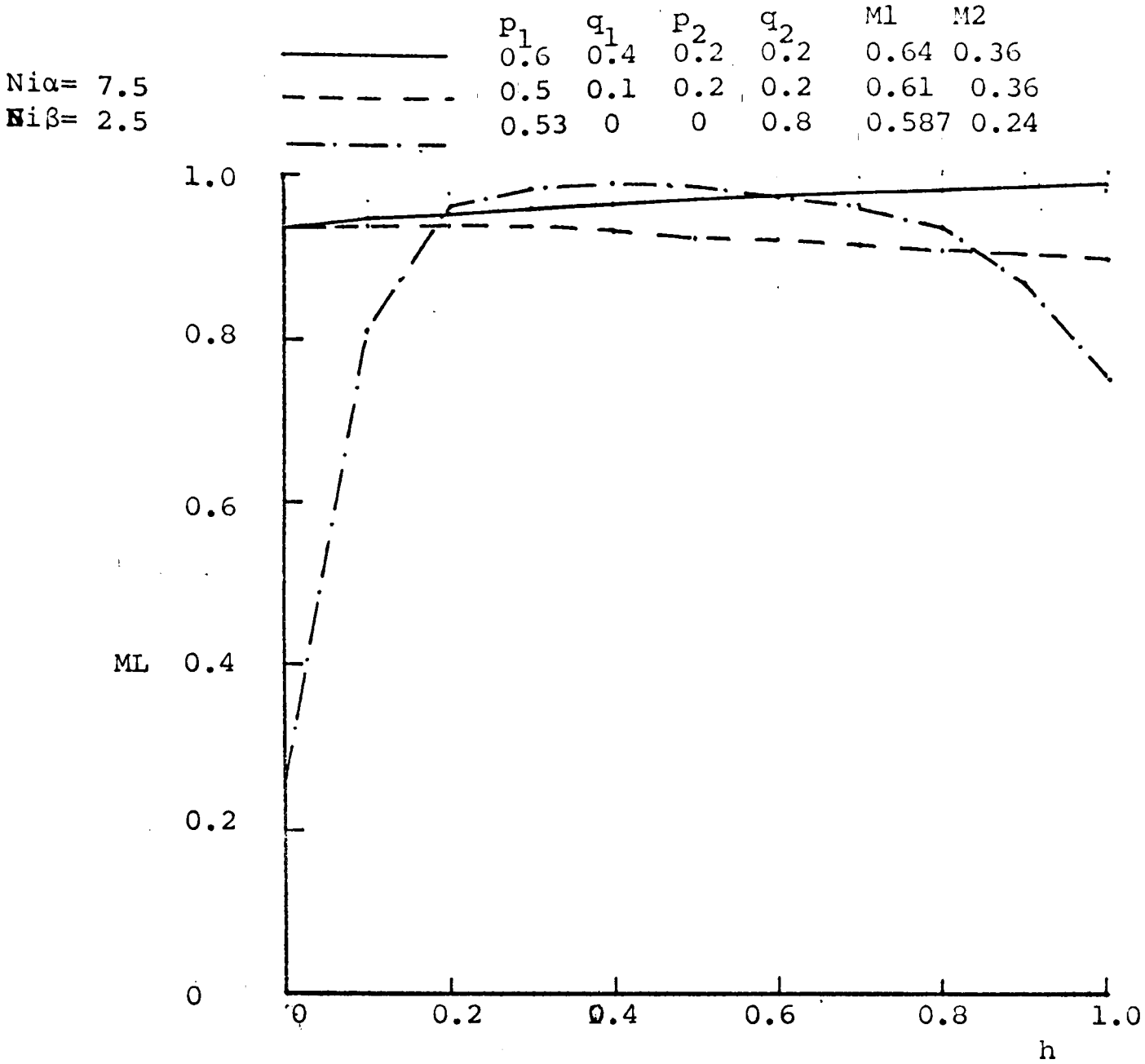


FIGURE 6.3 The relationship between the mean at the limit and h for independent recessive loci for N=10, n=2 under various models of initial frequency.



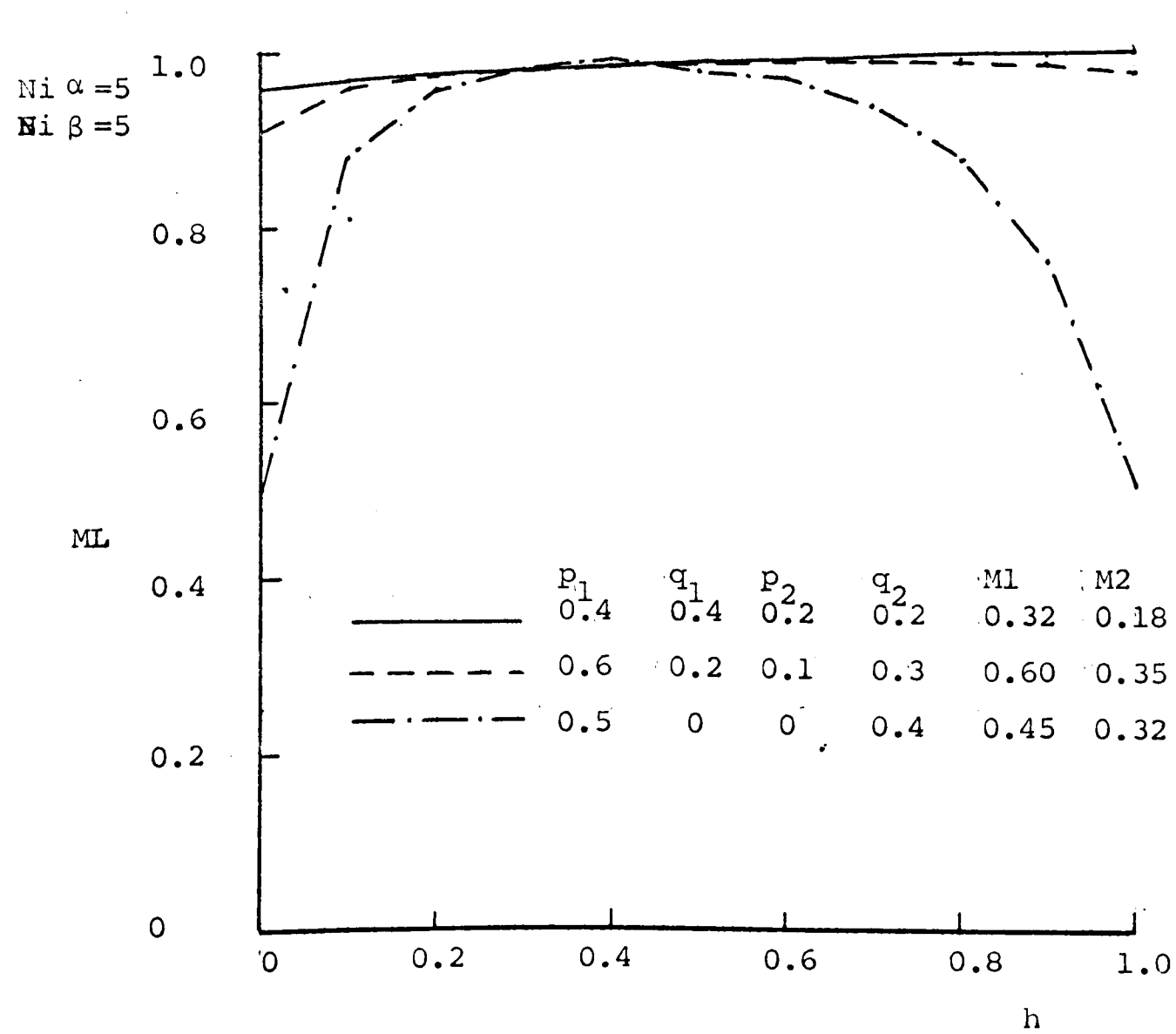


FIGURE 6.4. The relationship between the mean at the limit and h for independent dominant loci for N=10, n=2 under various models of initial frequency.

As in the previous chapter the next step is to examine the effect of linkage on at least some of the conclusions reached in this section. Initially a two locus model has been examined and then the multilocus case in the light of these results.

(c) Two Linked loci

Consider the situation where two linked loci are segregating in the two populations such that they are themselves in linkage equilibrium. Then let p_1, p_2, q_1 and q_2 be as previously defined, and let the cross be made by taking hN individuals from population 1 as before, then the disequilibrium in the cross is given by $D = h(1-h)(p_1-p_2)(q_1-q_2)$

$\therefore$ D is zero if $p_1=p_2$, or $q_1=q_2$ or h=1 or h=0.

For independent loci it has been shown that if gene action is additive there is no difference between

i) Immediate crossing followed by selection

ii) Selection in sub-lines followed by crossing and reselection, and that this holds for all values of h. This question is considered below for linked loci with:

i) No Recombination.

Firstly the case where Rc=0 will be considered, under this situation linkage disequilibrium has been shown to have its greatest effect. For this reason an extreme case with respect to linkage disequilibrium in the cross has been examined, that where $p_2=q_1=0$, then

$$D = h(1-h)p_1(-q_2) = -p_1q_2\ h(1-h)$$

For simplicity let $p_1=p$, $q_2=q$ then the gamete frequencies are

as follows

Gamete	Population 1	Population 2	Cross
AB	0	0	0
Ab	p	0	hp
aB	0	q	(1-h)q
ab	1-p	1-q	1-hp-(1-h)q
Mean	pα+K	qβ+K	hpα+(1-h)qβ+K

Let $u(\bar{p})$, $u(\bar{q})$, $v(p)$, $v(q)$, $u(p_1)$, $u(p_2)$, $u(q_1)$ and $u(q_2)$ be as previously defined.

Since only one locus is segregating in each population $u(p_1)$ etc can be given by single locus theory as

$$u(p_1) = \frac{1-e^{-2Ni\alpha hp}}{1-e^{-2Ni\alpha h}} \qquad u(p_2) = 0$$

$$u(q_1) = 0 \qquad u(q_2) = \frac{1-e^{-2Ni\beta(1-h)q}}{1-e^{-2Ni\beta(1-h)}}$$

and

$$v(p) = u(p_1)(1-u(q_2))\left[\frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}}\right] + u(p_1)u(q_2)\left[\frac{1-e^{-2Nih(\alpha-\beta)}}{1-e^{-2Ni(\alpha-\beta)}}\right]$$

$$v(q) = u(q_2)(1-u(p_1))\left[\frac{1-e^{-2Ni\beta(1-h)}}{1-e^{-2Ni\beta}}\right] + u(q_2)\left[u(p_1)\frac{1-e^{-2Ni(1-h)(\alpha-\beta)}}{1-e^{-2Ni(\beta-\alpha)}}\right]$$

...(24)

In general $u(\bar{p})$ and $u(\bar{q})$ can only be obtained by simulation but if $\alpha=\beta$ then the situation in the cross becomes effectively a two allele situation with $u(\bar{p})$ and $u(\bar{q})$ given by

$$u(\bar{p}) + u(\bar{q}) = \frac{1-e^{-2Ni\alpha(ph + (1-h)q)}}{1-e^{-2Ni\alpha}}$$

Some considerable algebraic manipulation of equations (74) above for $\alpha=\beta$ gives

$$v(p) + v(q) = \frac{1-e^{-2Ni\alpha(ph + (1-h)q)}}{1-e^{-2Ni\alpha}}$$

Therefore the value of ML under the sub-lining and crossing scheme is the same as that under the crossing and selecting scheme. However this is not in general true of the individual chance of fixation since

$$u(\bar{p}) = \frac{hp}{hp+(1-h)q}\left[\frac{1-e^{-2Ni\alpha(hp+(1-h)q)}}{1-e^{-2Ni\alpha}}\right] \quad \ldots(75)$$

while

$$v(p) = \frac{1-e^{-2Ni\alpha hp}}{1-e^{-2Ni\alpha h}}\left[\frac{e^{-2Ni\alpha(1-h)q} - e^{-2Ni\alpha(1-h)}}{1-e^{-2Ni\alpha(1-h)}} \times \frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}} + h\,\frac{1-e^{-2Ni\,(1-h)q}}{1-e^{-2Ni\,(1-h)}}\right] \quad \ldots(76)$$

for $h = 0.5$ we have

$$v(p) - u(\bar{p}) = \frac{(q-p)(1-e^{-Ni\alpha(p+q)}) + (p+q)(e^{-Ni\alpha q} - e^{-Ni\alpha p})}{2(p+q)(1-e^{-2Ni\alpha})} \quad \ldots(77)$$

$$= 0 \text{ if } p = q$$

Figure 6.5 shows the way in which $v(p)$, $u(\bar{p})$, $v(q)$ and $u(q)$ vary with q for $p=0.5$ and 0.1. $N=8$, $Ni\alpha=Ni\beta=4$, $Nc=0$, $h=0.5$. Under these conditions

$v(p) > u(\bar{p})$ for $q > p$ and $p \neq 0$

$v(p) < u(\bar{p})$ for $q < p$ and $q \neq 0$

and this conclusion holds over a wide variety of p, q and α values.

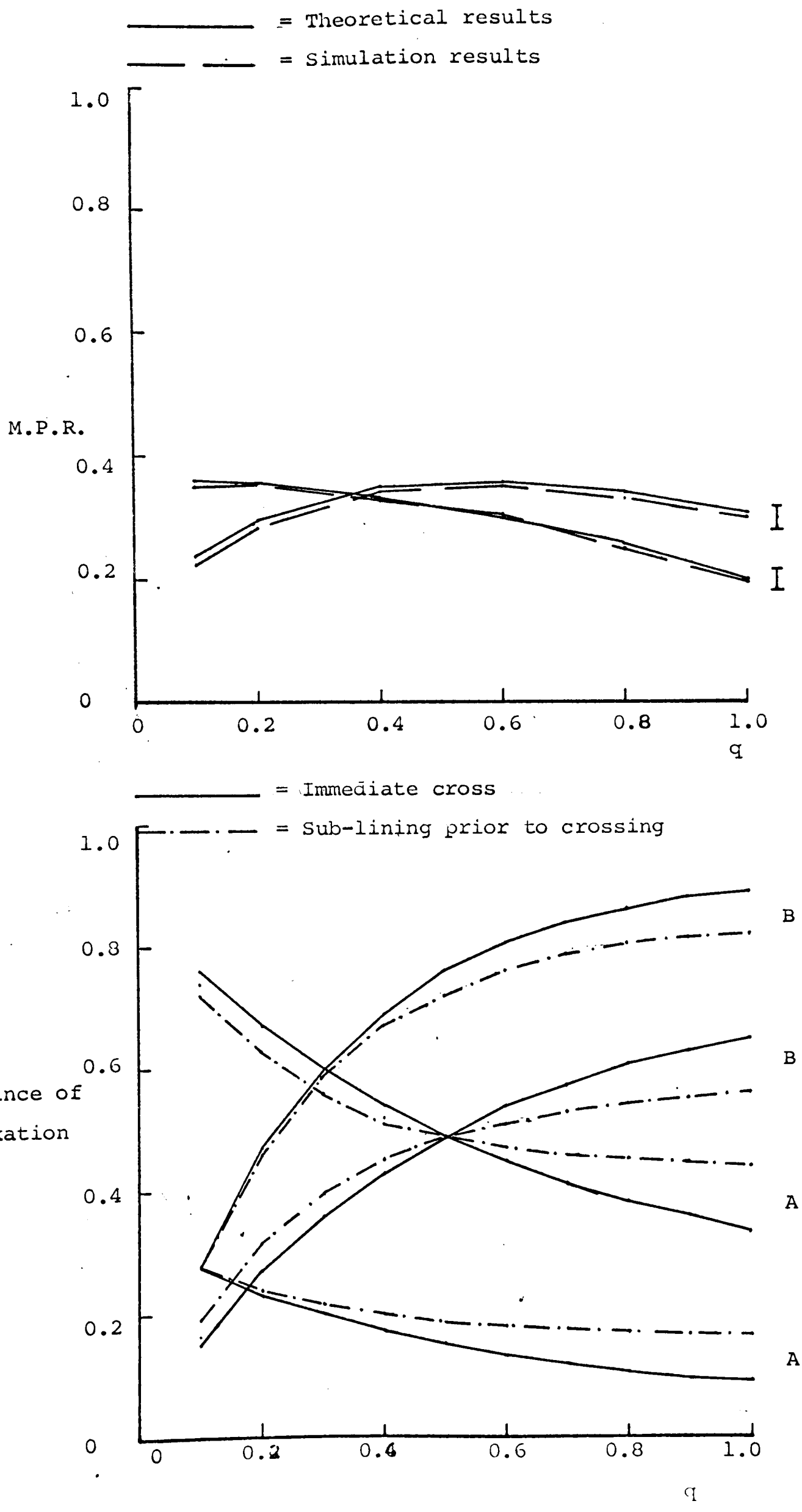


FIGURE 6.5. The effect of sub-line selection prior to crossing for N=8, n=2, Niα=Niβ=4, Nc=0, h=0.5 for various values of p and q with selection from two base populations. Theoretical results for the mean proportional response are compared with simulation with typical ranges of length four standard errors shown.

If $p = 0$ $v(p) = u(\bar{p}) = 0$

If $q = 0$ $v(p) = u(\bar{p}) = \dfrac{1-e^{-2Ni\alpha p}}{1-e^{-2Ni\alpha}}$

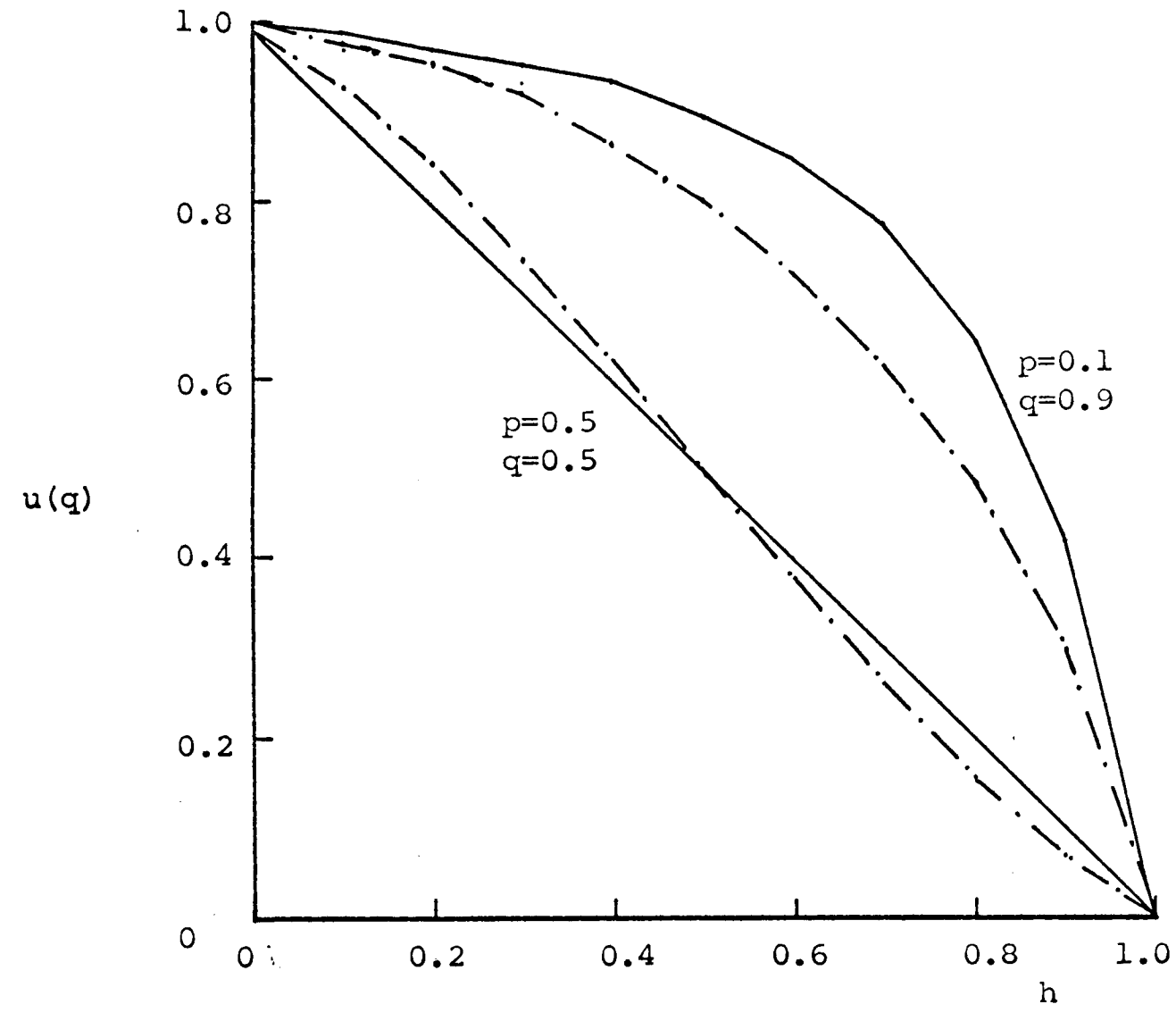
The way in which $v(p)$, $u(\bar{p})$, $v(q)$ and $u(\bar{q})$ vary with h is shown in Figure 6.6 for p=0.5, q=0.5 and p=0.1 q=0.9. This demonstrates that the difference in chance of fixation may vary considerably with h, reaching a maximum for some value of h which may be very different from 0.5

In general it has been found that:

(a) the lower frequency gene has a better chance of fixation under the sub-lining and crossing system and that this advantage is maximized for some value of h between 0.5 and 1 if $p < q$ and between 0 and 0.5 if $p > q$.

(b) the higher frequency gene has a better chance of fixation under the single line system. In all cases $v(p) - u(\bar{p}) = -(v(q) - u(q))$

For the case where $\alpha \neq \beta$, $u(\bar{p})$ and $u(\bar{q})$ can only be obtained in general by simulation. Since $v(p)$ and $v(q)$ can still be obtained algebraically there is a possibility that differences may appear due solely to the simulation procedure being introduced. To test for this, simulation was done for the case where $\alpha=\beta$ and compared with algebraic results, this is shown by Figure 6.5. The differences observed between the Mean Proportional Response values were statistically non-significant, although there was a tendency for simulation results to fall below the theoretical values by a very small amount. Typical results for the unequal effects case are shown in Figure 6.7 for N=8, $Ni\alpha=4$, $Ni\beta=2$, Nc=0, h=0.5, p=0.1 and 0.4 with chance of



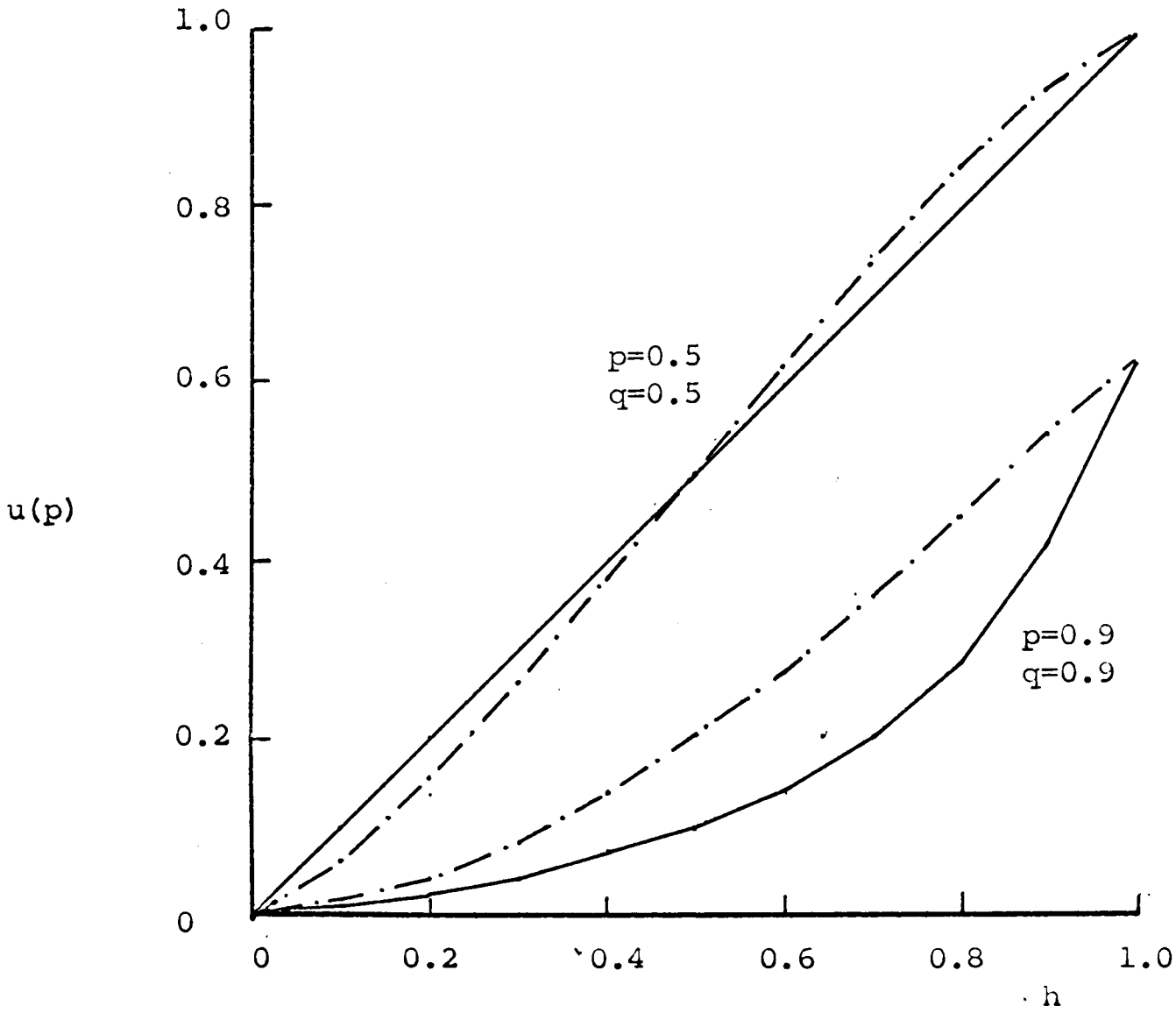
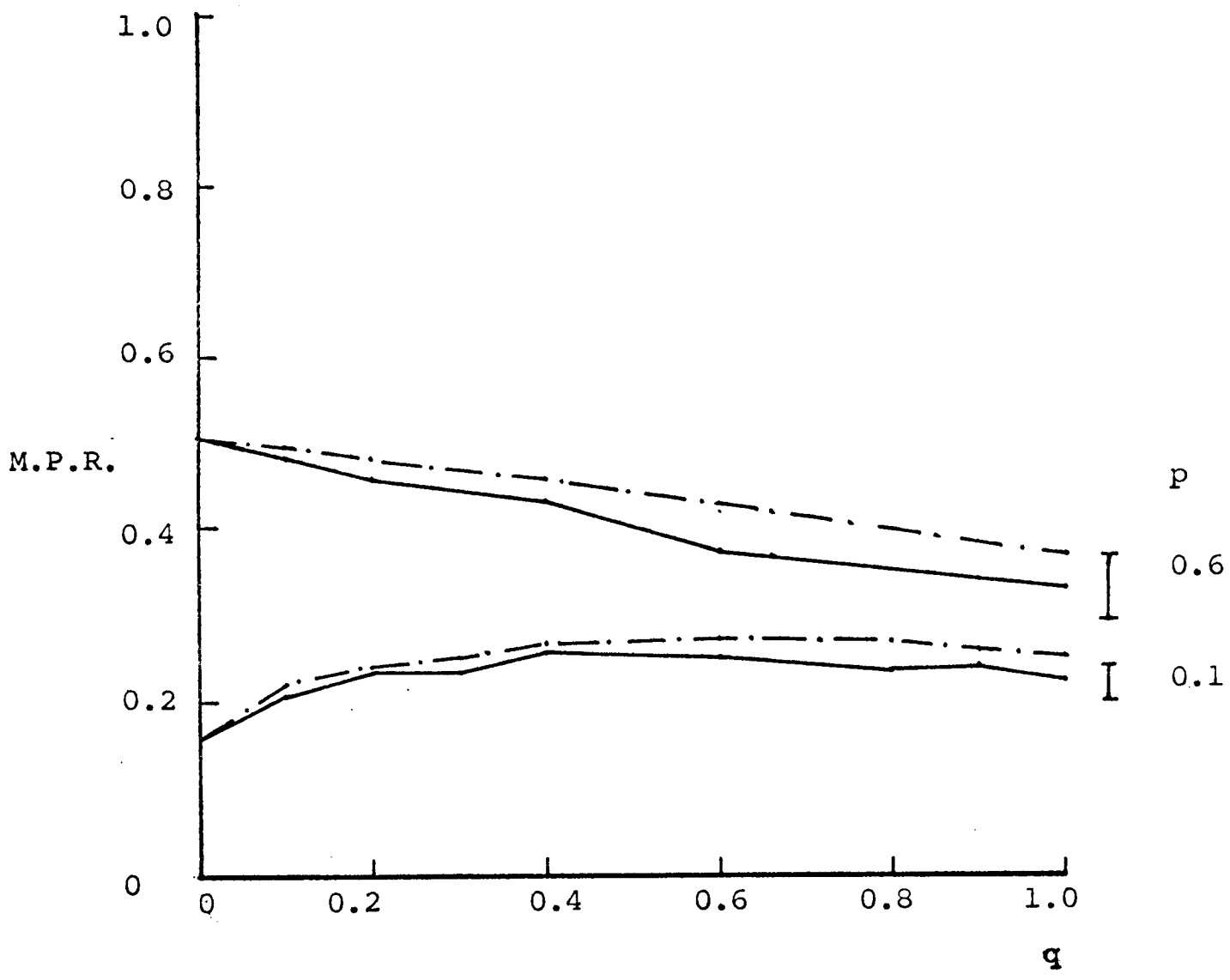


FIGURE 6.6. The relationship between chance of fixation of favourable alleles and h for N=10, n=2, Niα=Niβ=5, Nc=0 with selection from two base populations.



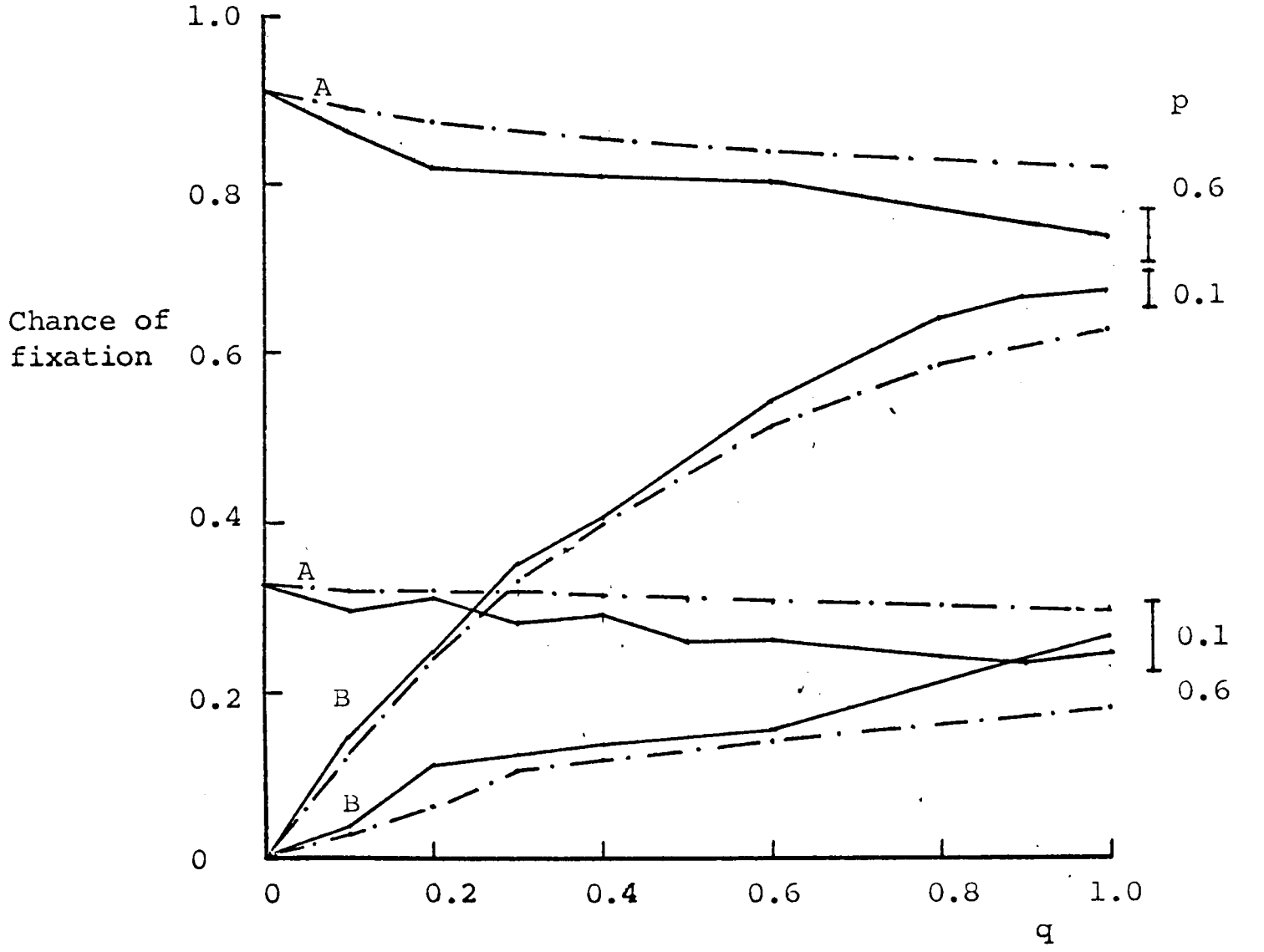


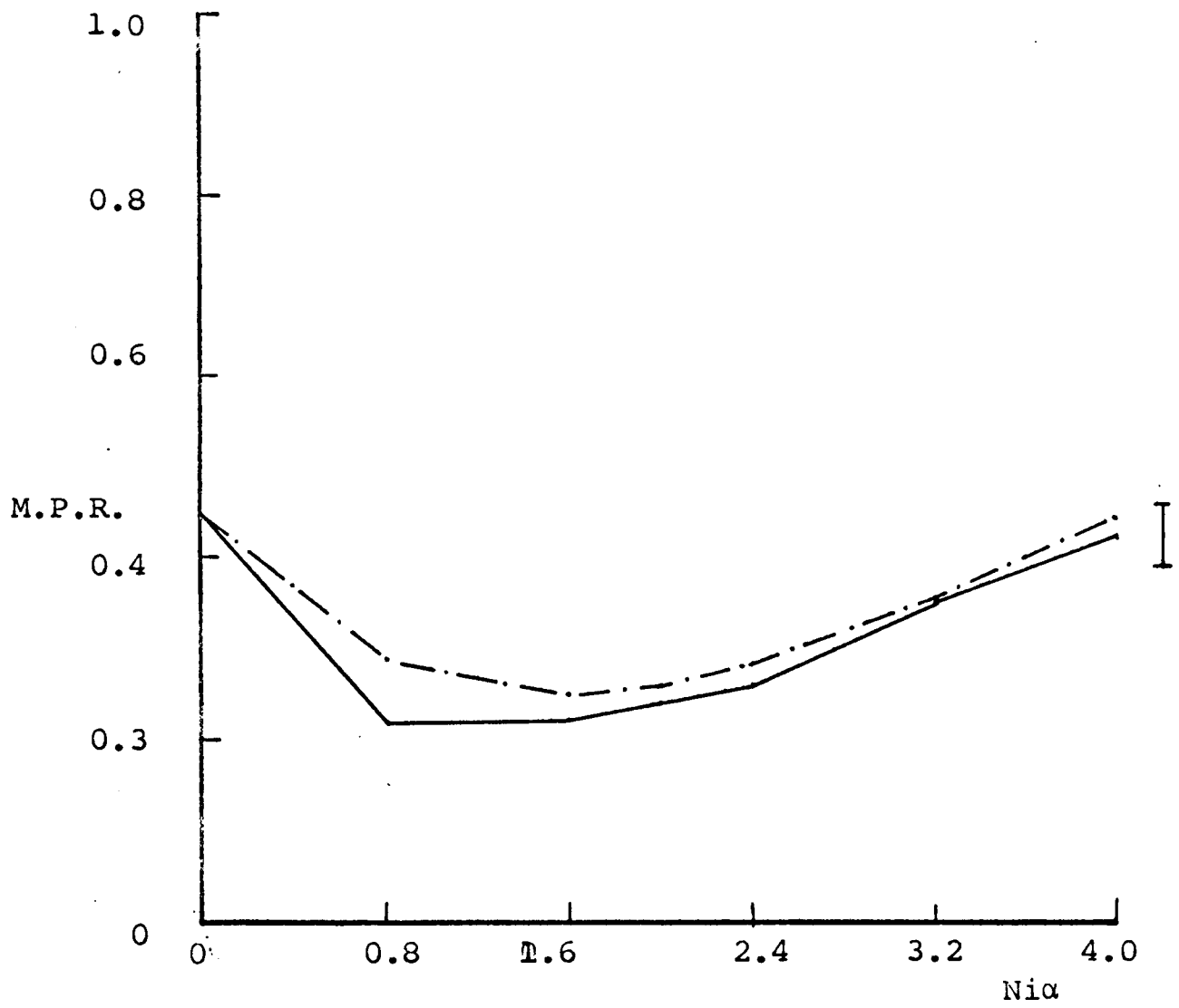
FIGURE 6.7. The effect of sub-line selection prior to crossing for N=8, n=2 Niα=4, Niβ=2, Nc=0, h=0.5 for various values of p and q with selection from two base populations. Typical ranges of length four standard errors are also shown.

fixation and M.P.R. plotted for a range of values of q. These results reveal that for the locus of larger effect, A, $v(p) > u(\bar{p})$ for all p and q examined, i.e. even with $p > q$. Conversely for the locus of smaller effect, B, $u(\bar{q}) > v(q)$ for all p and q, even when $p > q$. In general $v(p) - u(\bar{p}) \simeq - [v(q) - u(\bar{q})]$ and since the mean is mainly determined by the chance of fixation of the locus of larger effect, this means that the M.P.R. is greater under the sub-lining and crossing system. Figure 6.8 gives more results for the unequal effects case for $Ni\beta = 2$ and chance of fixation and M.P.R. plotted against $Ni\alpha$. Although obscured by sampling these results suggest that in general for p=q, of $\alpha > \beta$, then $v(p) > u(\bar{p})$ and $u(\bar{q}) > v(q)$ and if $\alpha < \beta$, then $v(p) < u(\bar{p})$ and $u(\bar{q}) < v(q)$.

This again gives the M.P.R. greater under the sub-lining system unless i)$\alpha = \beta$ or ii)$\alpha = 0$. In the latter case

$$v(q) = u(\bar{q}) = \frac{1-e^{-2Ni\beta(1-h)q}}{1-e^{-2Ni\beta}}$$

The differences observed between $u(\bar{p})$ and $v(p)$ for both the cases where $\alpha = \beta$ and $\alpha \neq \beta$ can be interpreted in terms of the effect of linkage and linkage disequilibrium. In the previous section on independent loci the result $v(p) = u(\bar{p})$ were found, therefore linkage must affect $v(p)$ and $u(\bar{p})$ differently in this situation. Since selection in the sub-lines involves only a single locus in each, linkage can have no effect in that part of the cycle, therefore it is only of importance in the cross situation. Let $v(p)_I$, $u(\bar{p})_I$ etc represent the chance of fixation for independent loci and $v(p)_L$, $u(p)_L$ etc. that for completely linked loci, then



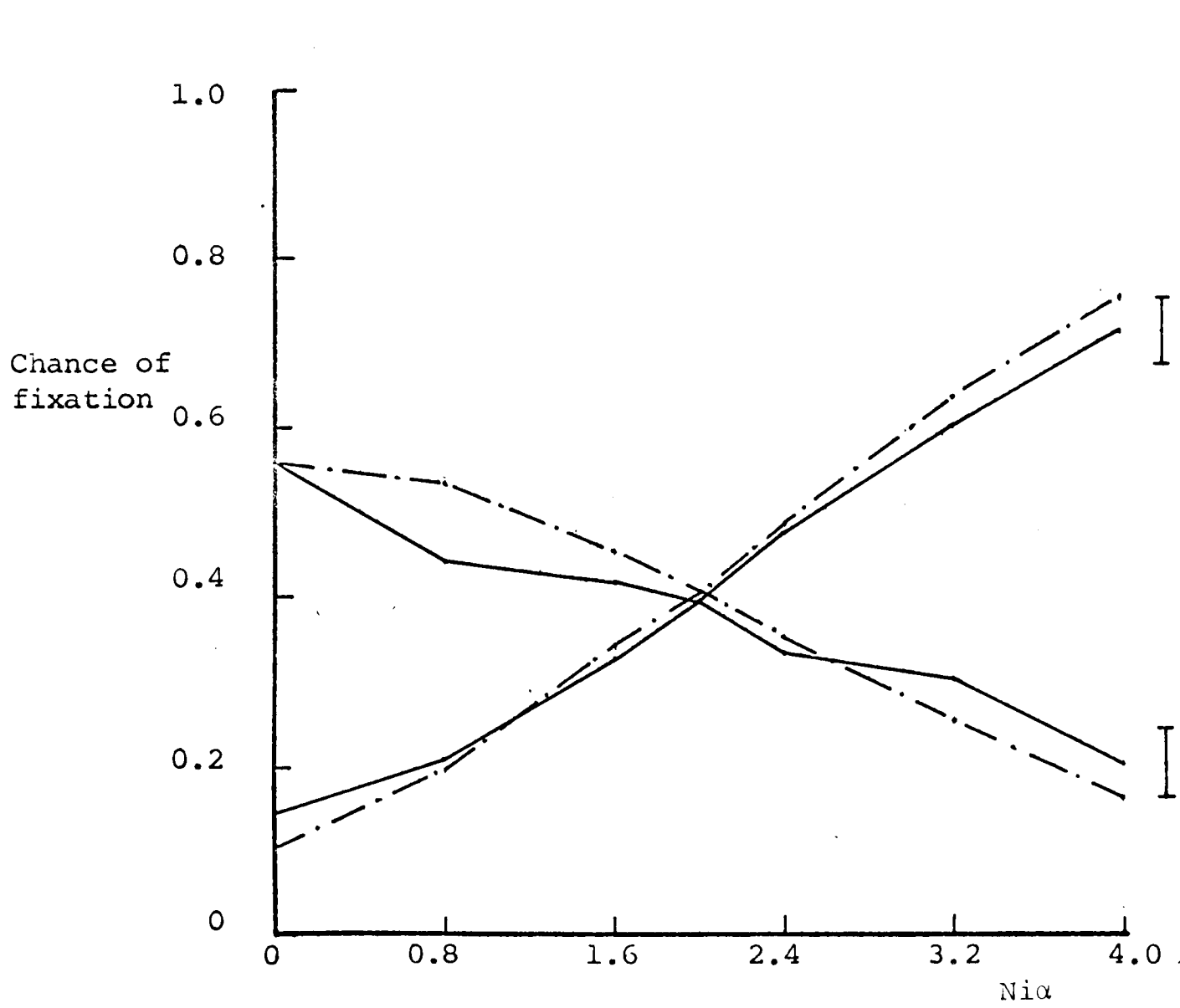


FIGURE 6.8 The influence of relative magnitude of gene effects on the differences between single line and sub-line selection systems for N=8, n=2, Niβ=2, Nc=0, p=q=0.4, h=0.5, with selection from two base populations. Typical ranges of length four standard errors are also shown.

$$u(\bar{p})_I = \frac{1-e^{-2Ni\alpha hp}}{1-e^{-2Ni\alpha}}$$

For $\alpha = \beta$

$$u(\bar{p})_L = \frac{hp}{hp+(1-h)q} \quad \frac{1-e^{-2Ni\alpha(hp+(1-h)q)}}{1-e^{-2Ni\alpha}}$$

$$\therefore u(p)_I - u(p)_L = \frac{hp(e^{-2Ni\alpha(hp+(1-h)q)} - e^{-2Ni\alpha hp}) + (1-h)q[1-e^{-2Ni\alpha hp}}{(hp+(1-h)q)(1-e^{-2Ni\alpha})} \quad \ldots(78)$$

$$v(p)_I = u(p_1)\left[\frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}}\right]$$

$$v(p)_L = u(p_1)\left[u(q_2)h + (1-u(q_2))\left[\frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}}\right]\right]$$

$$v(p)_I - v(p)_2 = u(p_1)u(q_2)\left[\frac{1-e^{-2Ni\alpha h}}{1-e^{-2Ni\alpha}} - h\right] \quad \ldots(79)$$

If the cross is made at the outset the effect of linkage is a function of the gene frequencies in the cross, i.e. hp and (1-h)q from equation (78). In general it has been found that the effect of linkage is greatest for the locus with its favourable allele at the lower frequency. If the cross is made after selection there are two possibilities.

i) Only one (or neither) locus segregating, in which case linkage has no effect.

ii) Both loci segregating with frequencies h and (1-h), then the effect of linkage is a function of these alone, from equation(79). If h = 0.5 the effect is the same for both loci, otherwise it is greater for the allele coming from the smaller sub-line.

Therefore, if h=0.5 and p≠q, the effect of linkage will be the same for each locus if the cross is made after the sub-line selection, but will be greater for the lower frequency allele if the cross is made at the outset. If hp=(1-h)q but h≠0.5 then the effect of linkage will be the same for each locus if the cross is made at the outset but will be greater for the locus coming from the smaller sub-line if the cross is made after sub-line selection.

When α≠β the chance of fixation for the immediate cross cannot be formulated but from an earlier section on the effect of linkage disequilibrium it is clear that in any cross which generates disequilibrium its effect will be greatest for the smaller allele. Selection in the sub-lines prior to making the cross will increase the disequilibrium generated in the cross, therefore the allele of smaller effect should have its chance of fixation reduced by sub-line selection. This might also be expected to apply to the allele with larger effect, although to a lesser extent, but since Nc=0 the following must be true

$$u(f_2)_L + u(f_3)_L + u(f_4)_L = 1$$

$$v(f_2)_L + v(f_3)_L = 1$$

so that if $\alpha > \beta$ and $v(f_3)_L < u(f_3)_L$

then $v(f_2)_L > u(f_2)_L$.

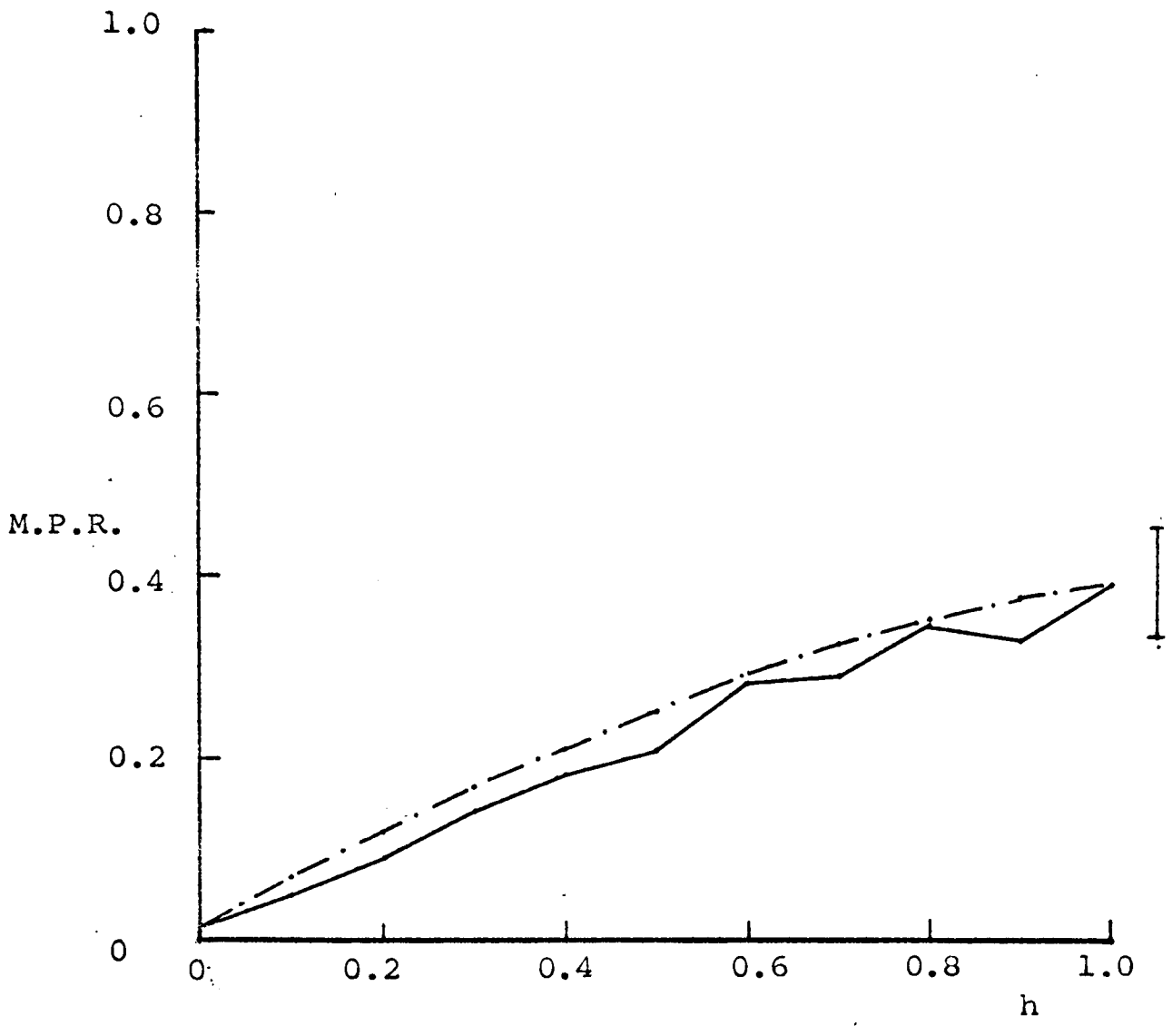
Therefore as sub-lining and crossing decreases the chance of fixation of the allele with smaller effect so it must increase the chance of fixation of the larger allele, If α and β are reasonably large then $u(f_4)_L \approx 0$ and $v(f_2)_L - u(f_2)_L \approx -(v(f_3)_L - u(f_3)_L)$ i.e. $v(p) - u(\bar{p}) \approx -(v(q) - u(\bar{q}))$ as was found.

From Figure 6.7 it appears that this is approximately true for a range of different p and q values, even when q is high and p is low and this is in agreement with earlier findings regarding the effect of linkage disequilibrium.

To see the effect of changing h under these conditions simulation was done for $\alpha > \beta$ with h varying from 0 to 1. Results are given in Figure 6.9, for all $h \neq 1.0$ the sub-lining system gave a higher M.P.R. without any indication of any particular value giving a larger difference.

The optimal value of h with respect to mean response is given by the value of h maximizing $\alpha v(p) + \beta v(q)$. If $\alpha=\beta$ this is given by one if $p > q$ or zero if $q > p$. If $\alpha > \beta$ and $p \geq q$ then $h = 1$ will give the maximum response but if $p < q$ then an intermediate value of h may be advisable.

Consider next the rather more general situation when p_2 and q_1 are not zero and how this affects conclusions from the special case discussed above. As $p_1 \to p_2$ and $q_1 \to q_2$ so $D \to 0$ and the single base population is approached. To see the effect of this $\bar{p}$ and $\bar{q}$ were kept constant and p_1, p_2, q_1 and q_2 varied such that $\bar{p} = 2\bar{q}$ and $p_1 = 2q_2$ so that D varied from 0 to $-\frac{\bar{p}\bar{q}}{4}$ for h=0.5. Figure 6.10 gives chance of fixation and M.P.R. plotted against D for $\bar{p} = 0.5$, $\bar{q}=0.25$. This shows that as D approaches zero so the effect of the sub-line selection is rapidly reduced. As $\alpha=\beta$ in this case there are no differences in the M.P.R. for any D value. The case for $\alpha \neq \beta$ is shown in Figure 6.11 where the M.P.R. is greater under the sub-lining system only for large negative value of D.



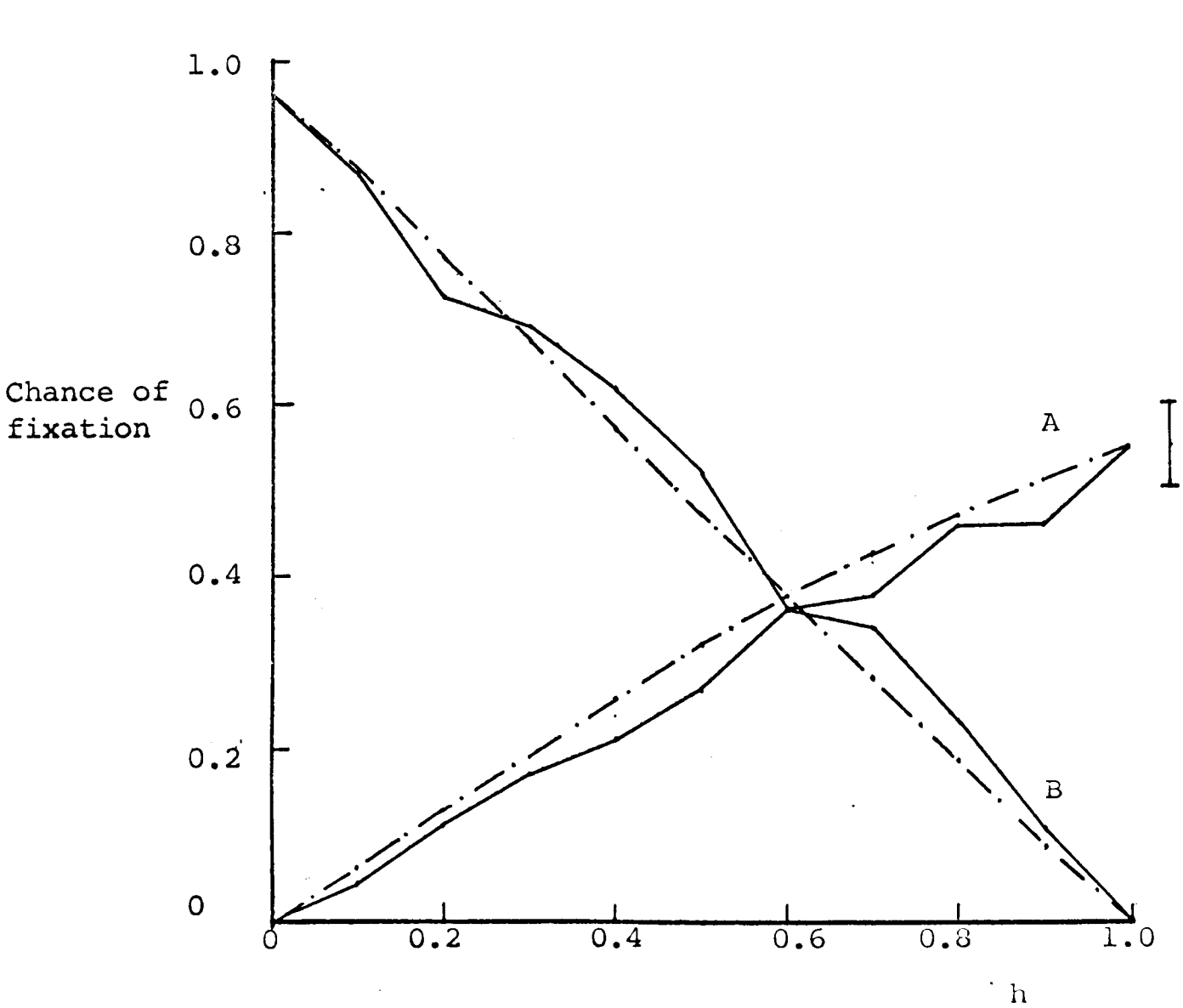
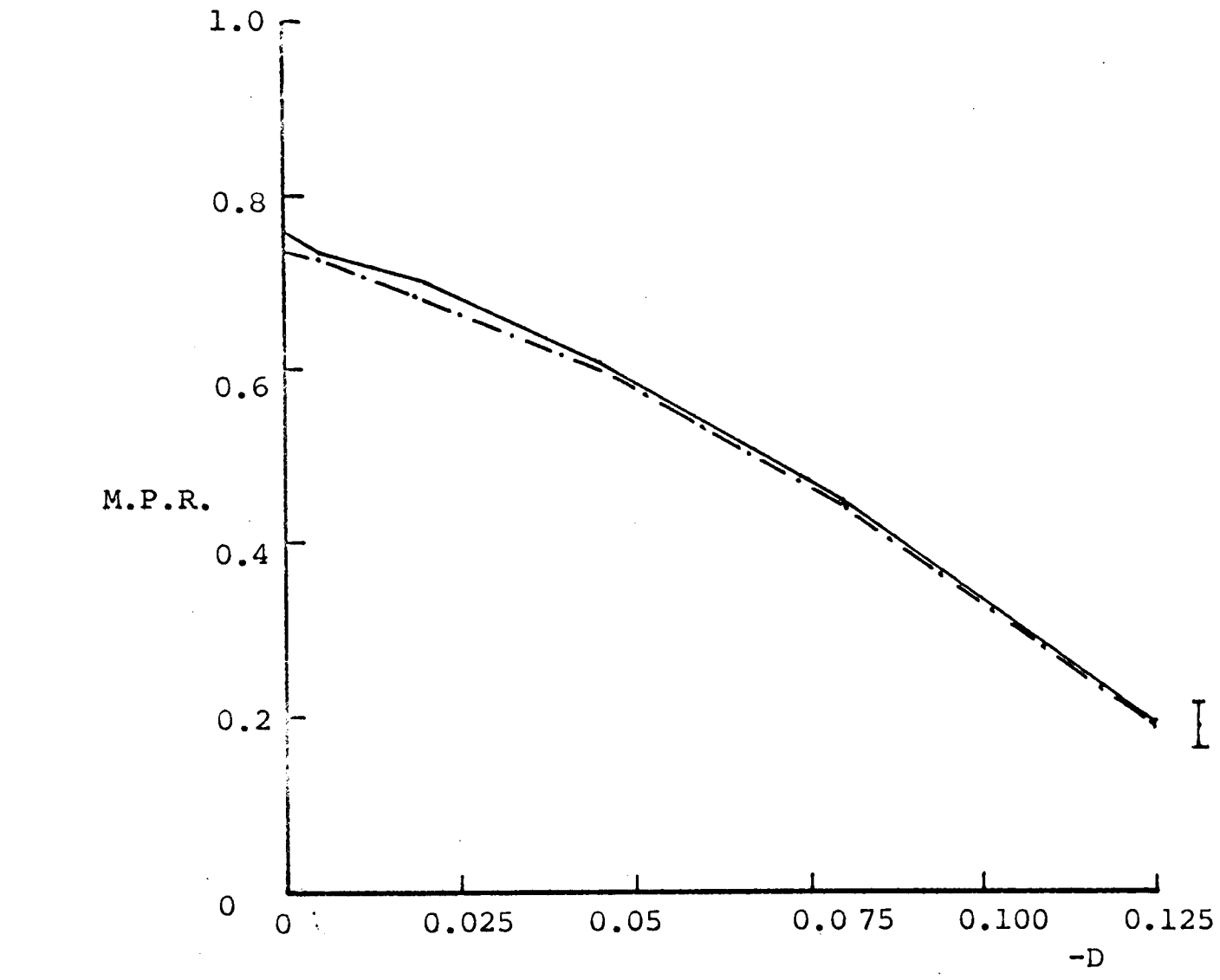


FIGURE 6.9. The influence of h on the effect of sub-line selection prior to crossing for N=8, n=2, Niα=4, Niβ=1, Nc=0, p=0.1, q=0.9 with selection from two base populations. Typical ranges of length four standard errors are also shown.



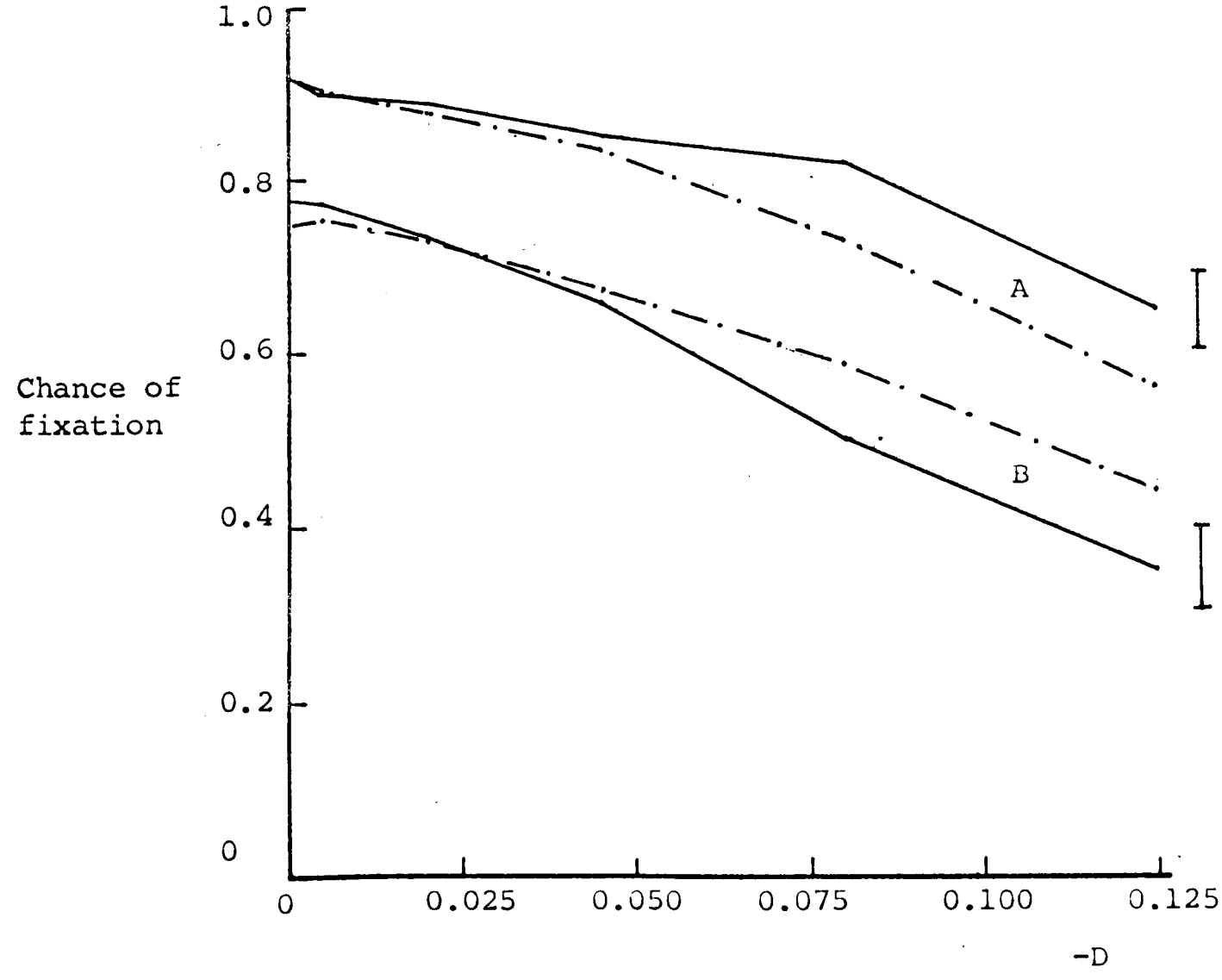
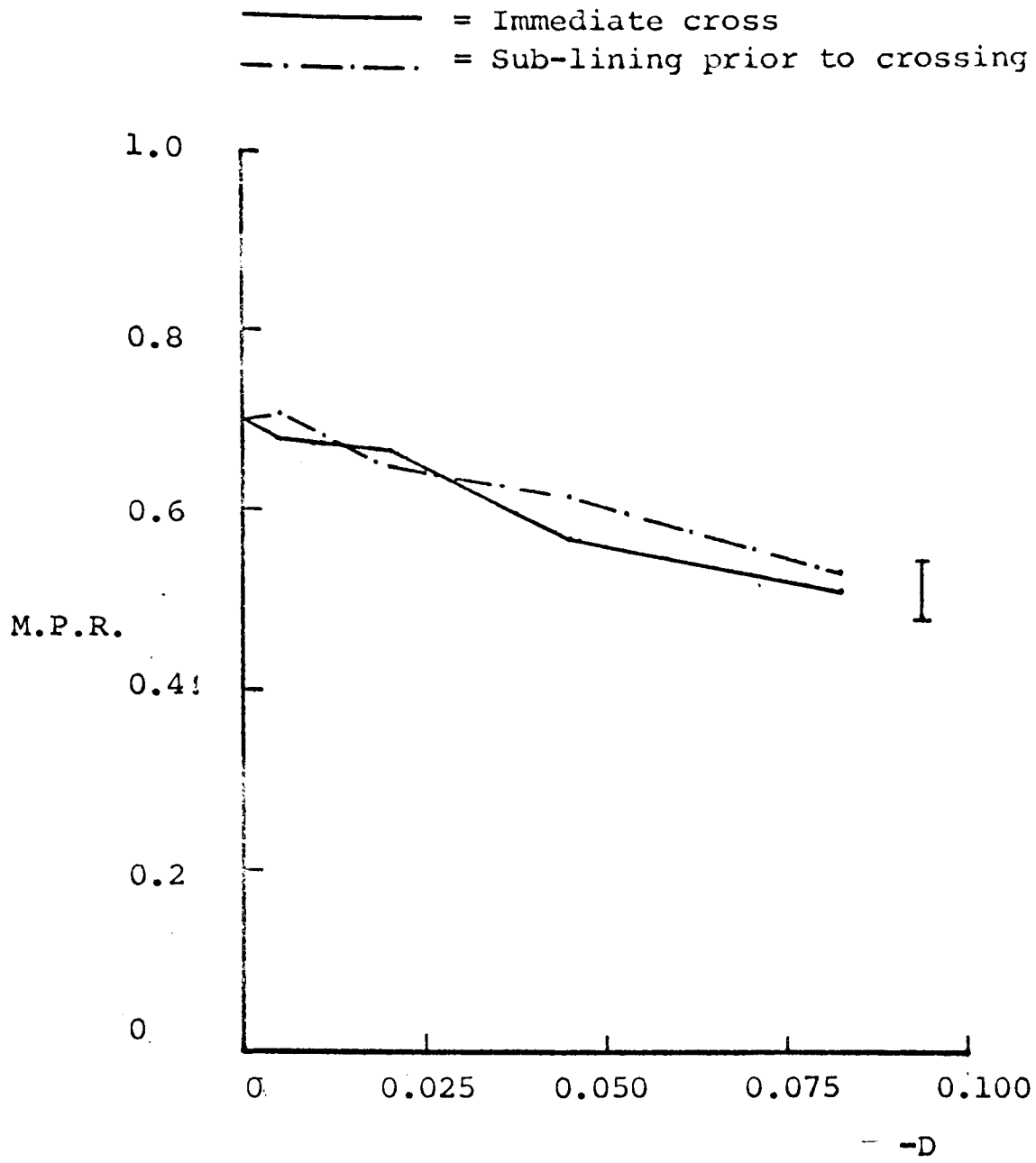


FIGURE 6.10. The influence of initial linkage disequilibrium in the cross for the same mean frequencies on the effect of sub-line Selections prior to crossing, for N=8, n=2, Niα=Niβ=4, Nc=0, h=0.5, $p=p_1+p_2=1.0$, $q=q_1+q_2=0.5$, $p_1=2q_2$, $p_2=2q_1$, $D_o=\frac{(p_1-p_2)(q_1-q_2)}{4}$ with selection from two base populations. Typical ranges of length four standard errors are also shown.



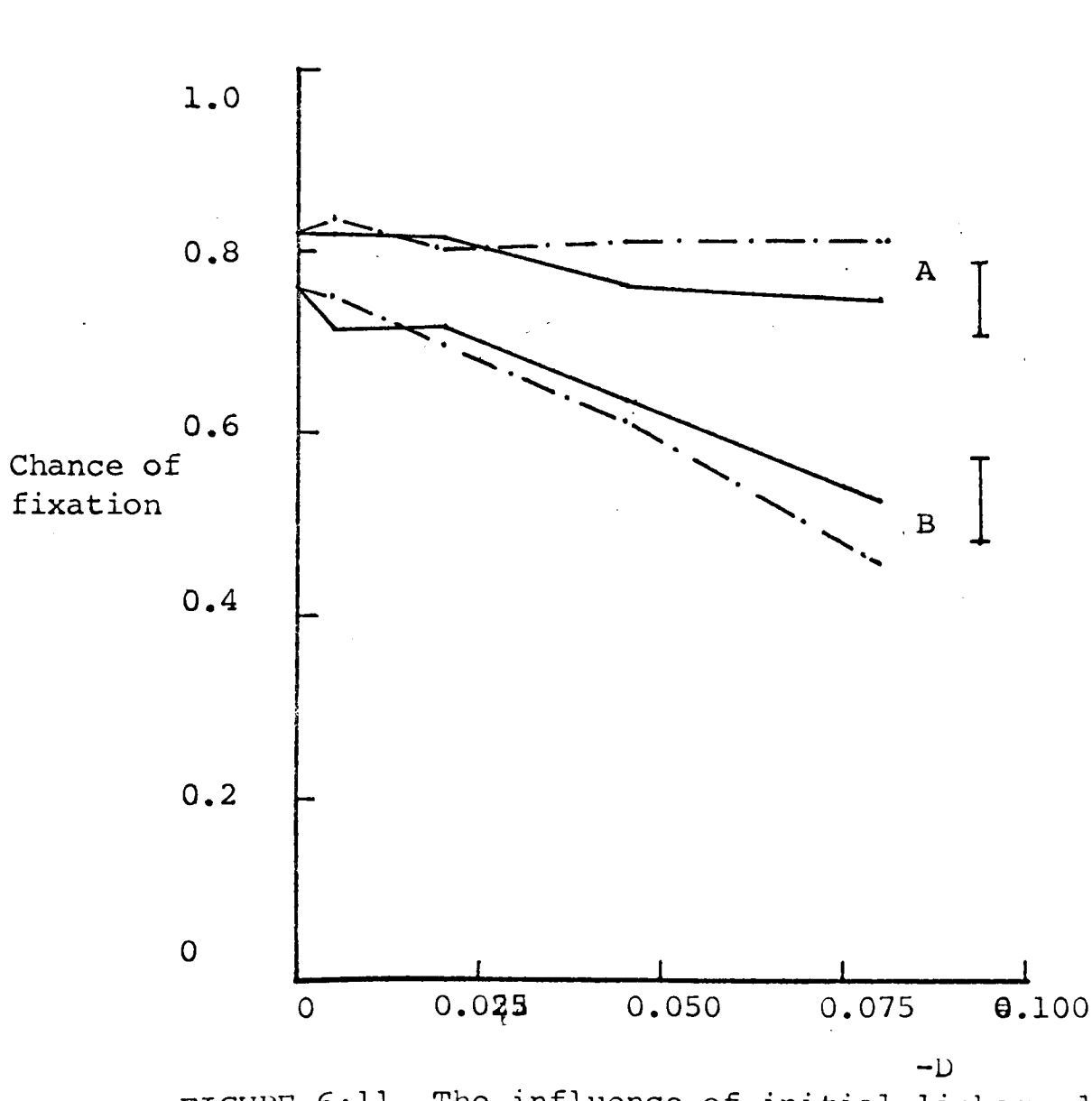
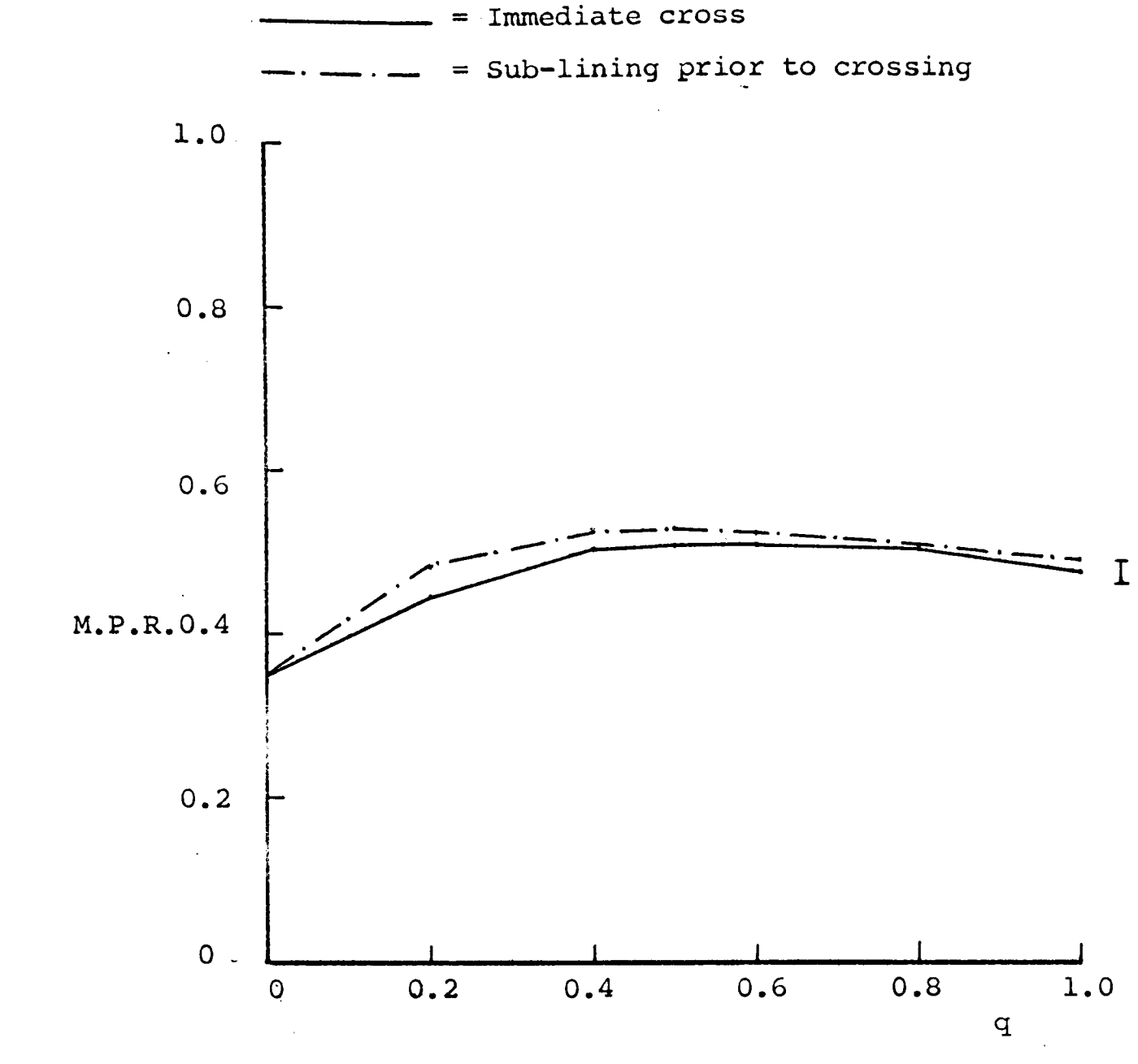


FIGURE 6;11. The influence of initial linkage disequilibrium in the cross for the same mean frequencies on the effect of sub-line selection prior to crossing for N=8, n=2, Niα=4, Niβ=2, Nc=0, h=0.5, $p=p_1+p_2=0.5$ $q=q_1+q_2=1$, $q_1=2p_2$, $q_2=2p_1$, $D_o=\frac{(p_1-p_2)(q_1-q_2)}{4}$ with selection from two base populations. Typical ranges of length four standard errors are also shown.

ii) Recombination

Returning to the extreme case where $p_2=q_1=0$ and $D= -h(1-h)p_1q_2$, consider the consequences of introducing intermediate recombination. This will have no effect on the chance of fixation in the sub-lines, nor in any cross involving only one (or neither) segregating locus. In those crosses where both loci segregate the chance of fixation cannot be formulated and simulation must be used. Some results for $Nc=\frac{1}{4}$ are presented in Figures 6.12 and 6.13 for $\alpha=\beta$, comparison with 6.5 reveals that recombination of an intermediate nature has little effect on the difference between crossing at the outset, or selection prior to crossing for the lower frequency allele. However the difference for the higher frequency allele is drastically reduced and frequently changed in sign so that it too gains from the sub-line selection. The net result is to give in all cases a significantly higher M.P.R. under the sub-line selection system. This implies that some intermediate amount of recombination is of more value to the cross if it is made after sub-line selection, which is reasonable since the disequilibrium generated in that case is of a greater magnitude. Figure 6.14 gives results for $\alpha = \beta$, $q > p$ with chance of fixation and M.P.R. plotted against $Nc/2(1+2Nc)$ again showing that sub-line selection gives a significantly higher M.P.R. although differences are still quite small, being of the order of 0.05.

With unequal effects and recombination it is no longer true that if $v(q) < u(\bar{q})$, then $v(p) > u(\bar{p})$ since $u(\bar{q}) + u(\bar{p})$ may then be greater than 1. Figure 6.15 gives results for



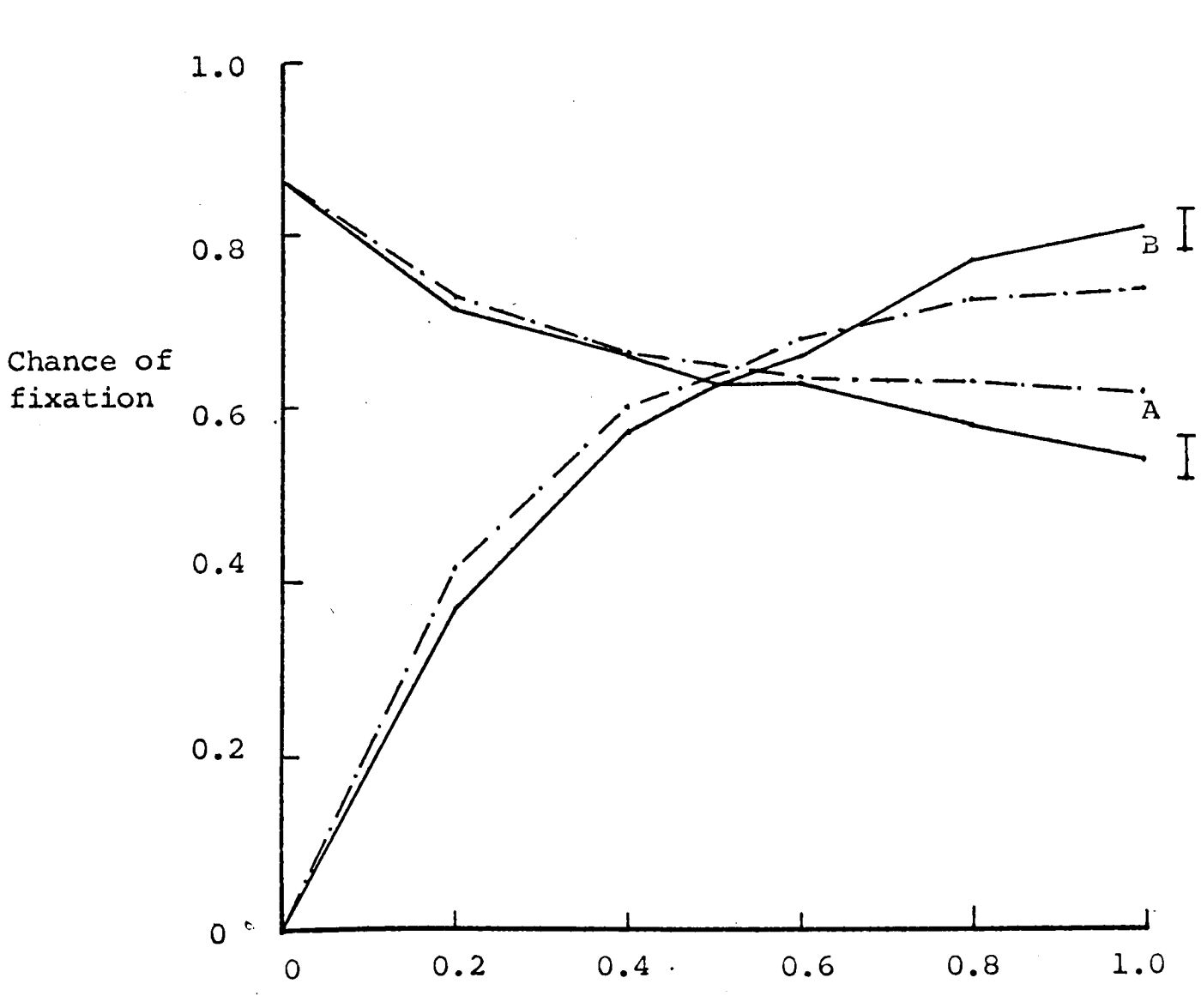


FIGURE 6.12. The effect of sub-line selection prior to crossing for N=8, n=2, Niα=Niβ=4, Nc=0.25, p=0.5, n=0.5 with selection from two base populations. Typical ranges of length four standard errors are also shown .

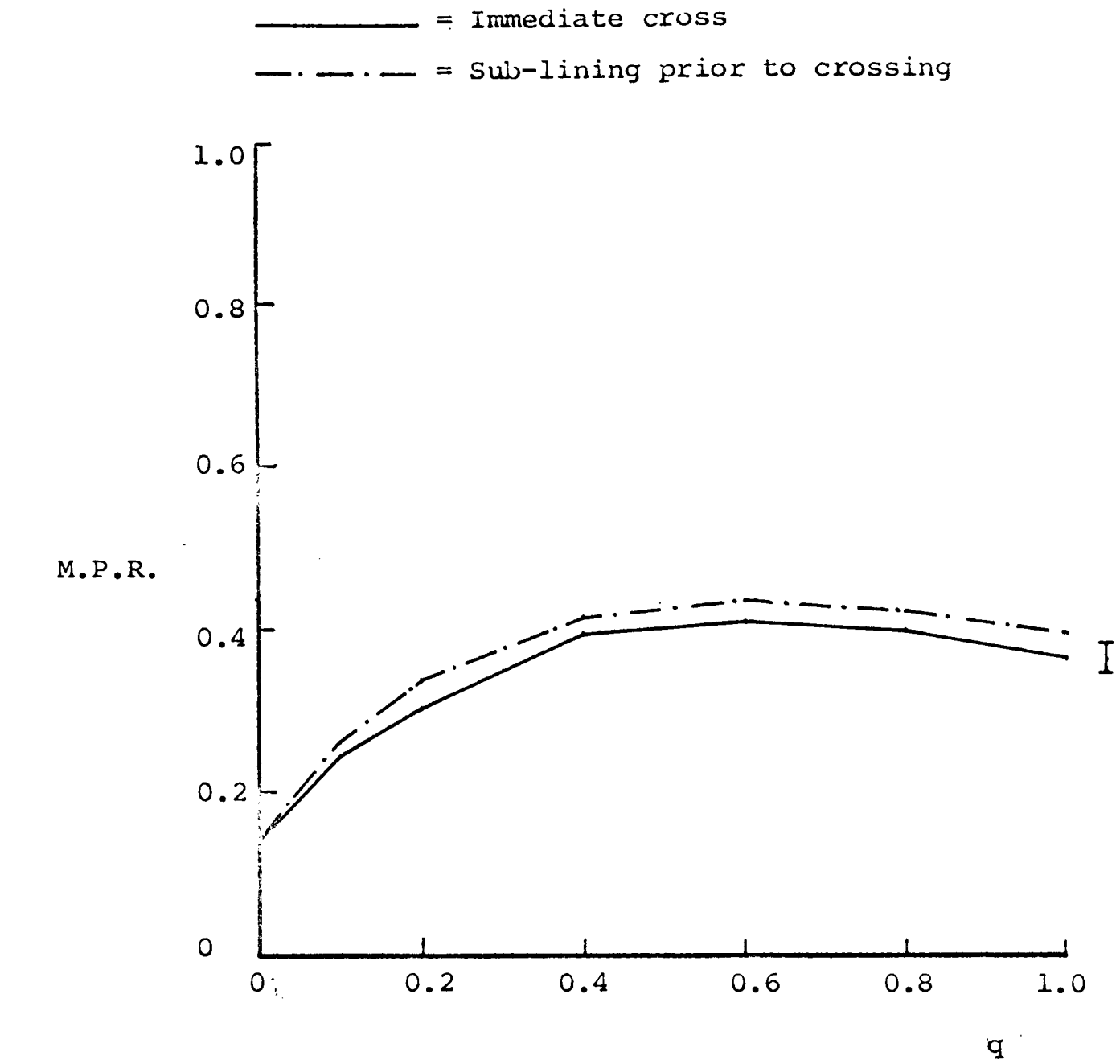


FIGURE 6.13. The effect of sub-line selection prior to crossing for N=8, n=2, Niα=Niβ=4, Nc=0.25, p=0.10, h=0.5 with selection from two base populations. Typical ranges of length four standard errors are also shown.

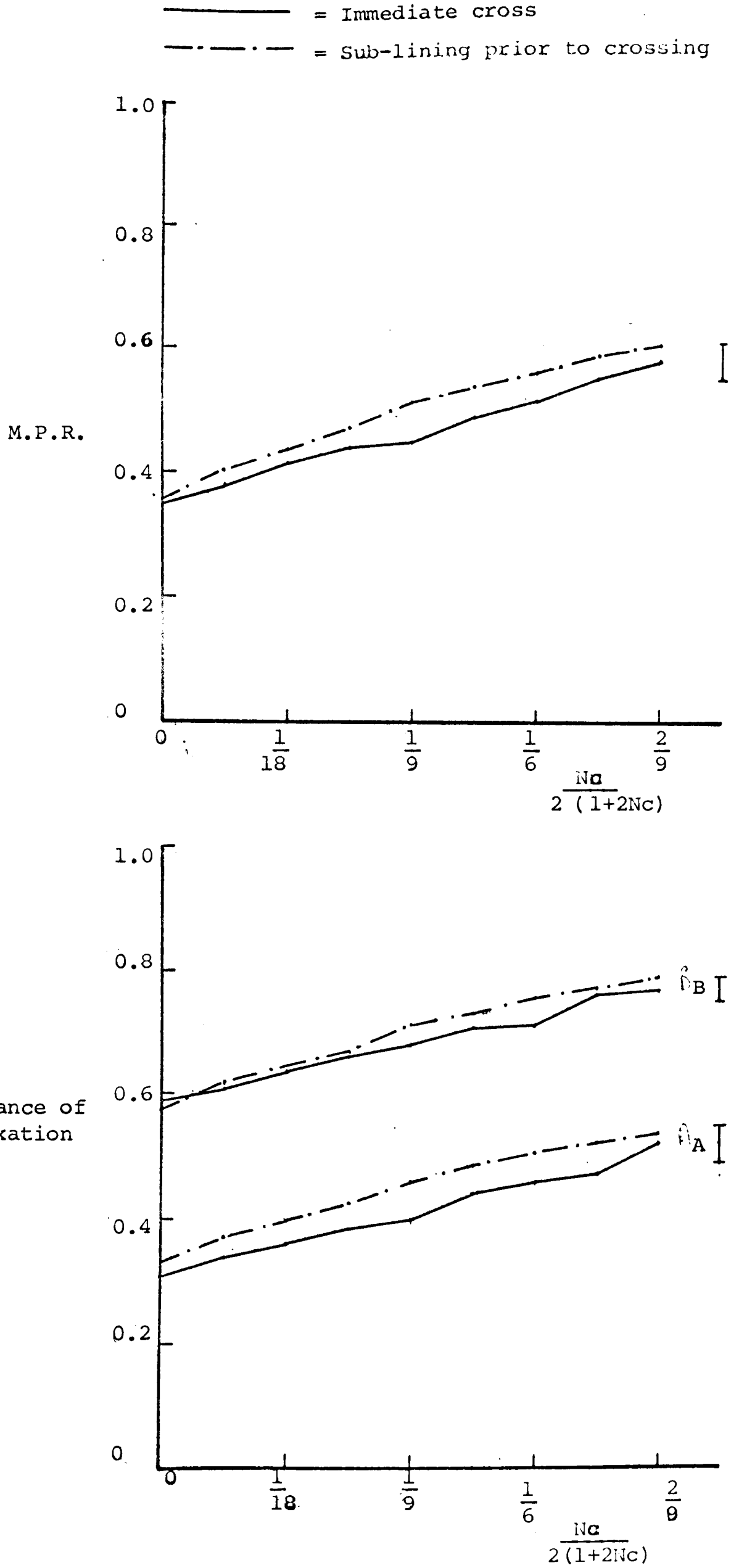


FIGURE 6.14. The influence of recombination fraction on the effect of sub-line selection prior to crossing for N=8, Niα=Niβ=4, p=0.1, q=0.2, h=0.5 with selection from two base populations. Typical ranges of length four standard errors are also shown.

$Ni\beta=2$ with variable $Ni\alpha$ showing that with recombination and equal frequencies there is no longer any clear advantage to the sub-lining system.

Another aspect of recombination concerns the choice of h value. Consider for example the case where $\alpha=\beta$ and $p=q$:

If $Nc=0$ $u(\bar{p}) + u(\bar{q}) = v(p) + v(q) = \dfrac{1-e^{-2Ni\alpha p}}{1-e^{-2Ni\alpha}}$

for all h.

If $Nc = \infty$ $u(\bar{p}) + u(\bar{q}) = v(p) + v(q) = \dfrac{2 - e^{-2Ni\alpha hp} - e^{-2Ni\alpha(1-h)p}}{1-e^{-2Ni\alpha}}$

which is maximized for $h = 0.5$.

Therefore recombination favours the use of an intermediate h value.

If $\alpha=\beta$ but $p > q$ then

If $Nc = 0$ $u(\bar{p}) + u(\bar{q})$ is maximized for $h = 1$

If $Nc = \infty$ $u(\bar{p}) + u(\bar{q})$ is maximized for $h = \dfrac{\log_e(p/q)}{2Ni\alpha} + \dfrac{q}{p+q}$

If $\alpha\neq\beta$ the value of h which maximizes response cannot easily be obtained even for $Nc=\infty$ but in general some intermediate value of h will give the greatest response if different loci segregate in the two populations and if there is some recombination.

These studies have revealed two situations under which sub-line selection prior to crossing might give a higher response, these are:

(a) No recombination and loci of unequal effect.

(b) Intermediate recombination and loci of equal effect.

The next section examines the validity of these conclusions under a multilocus model and also considers the consequences of practising sub-ling selection only for a limited period.

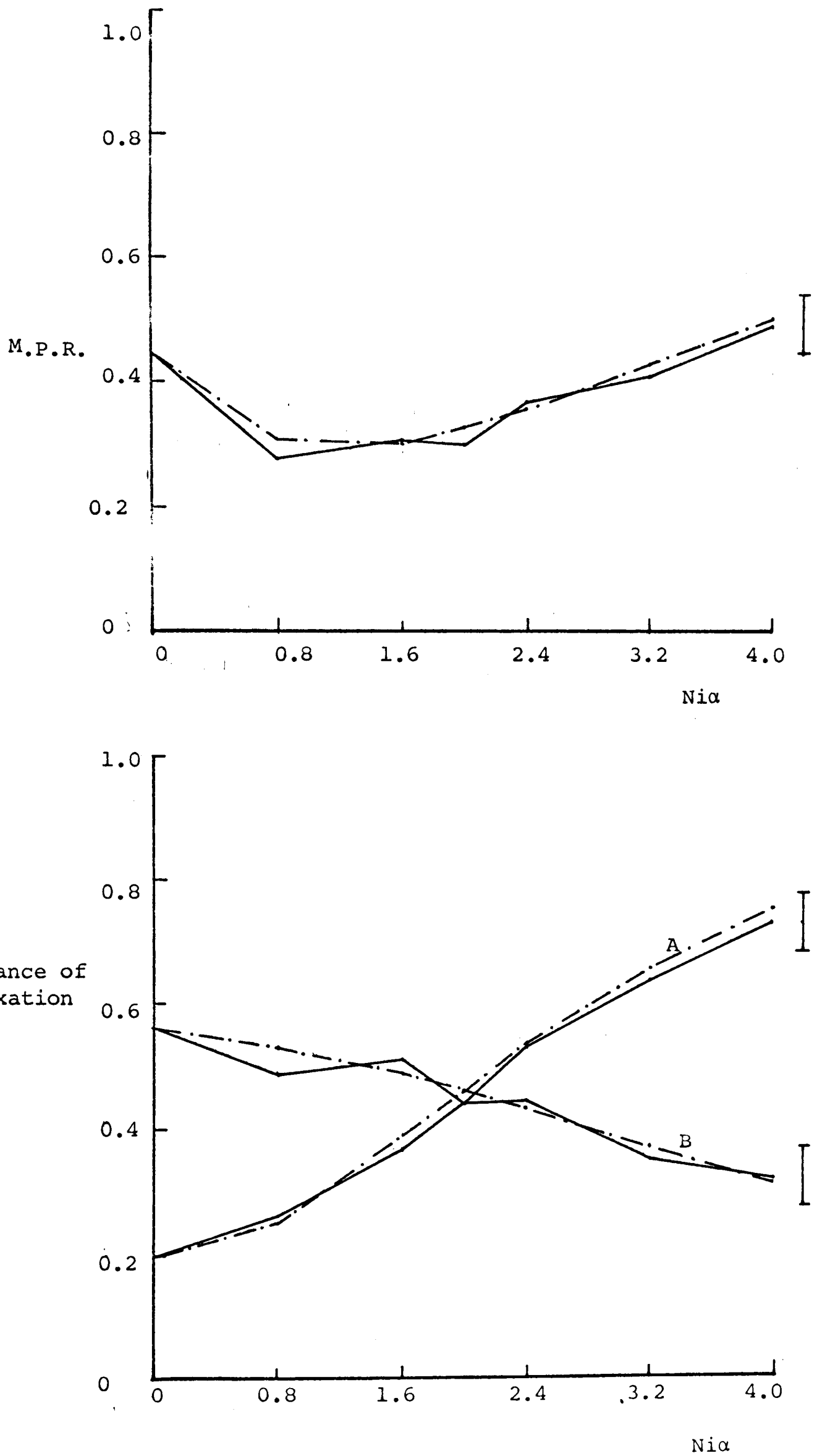


FIGURE 6.15. The influence of relative magnitude of gene effects on the difference between single line and sub-line selection systems for N=8, n=2, Niβ=2, p=q=0.4, h=0.5 with selection from two base populations. Typical ranges of length four standard errors are also shown.

(d) Multi-locus Models

The earlier chapter in the effects of disequilibrium revealed that as the number of loci, n, increases so the maximum amount of negative disequilibrium which can be generated is decreased. Even for n=4 it was found that the effect of disequilibrium, as produced by making a cross between distinct populations, could be quite considerably reduced from the two locus case. Therefore it is to be expected that the differences found between the immediate cross and sub-line selection systems for the two locus model will be reduced under a multi-locus model since they are associated with the presence of negative linkage disequilibrium.

A four locus model has again been studied using simulation but since the range of possible parameter sets is vast nothing like a full investigation has been undertaken. Instead some special cases have been examined in particular those which the two locus studies suggest might give differences between the sub-line selection system and the immediate cross system.

The first question to be considered concerns the number of generations, T, for which sub-line selection is practised. This has been examined for a case which gave differences under the two locus model and is described below:

Chromosome	A	B	C	D
Effects	a	a	a	a
Recombination fraction	c	c	c	
Frequency in $Pop^n.1$	q_1	q_1	0	0
Frequency in $Pop^n.2$	0	0	q_2	q_2

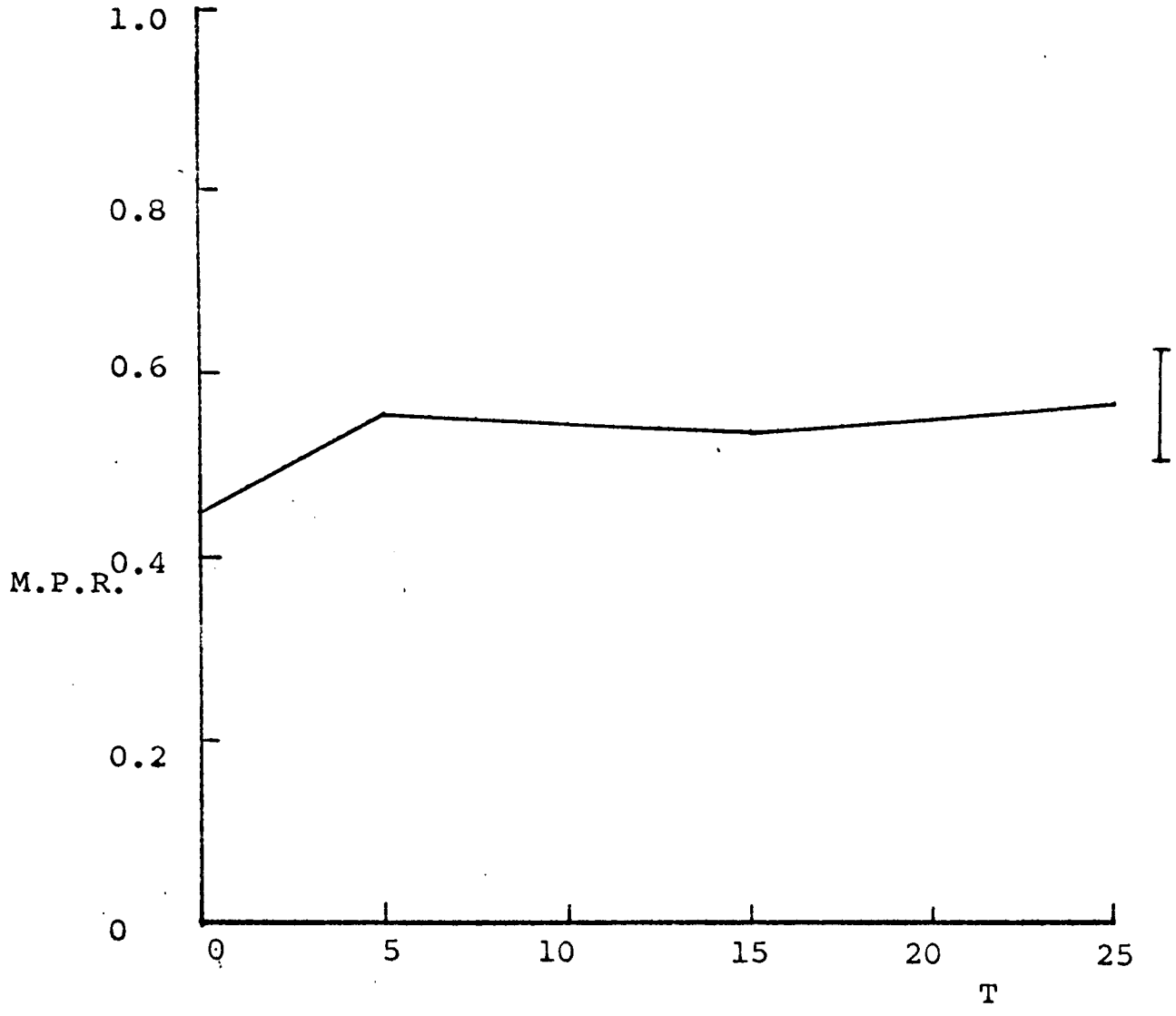
such that $c > 0$ and $q_1 \neq q_2$.

Results for $q_1 = 0.4$, $q_2 = 0.8$, $Nc = \frac{1}{4}$ are given in Figure 6.16 with chance of fixation and M.P.R. plotted against T. This shows that as T increases so the chance of fixation of the lower frequency allele tends to increase while that of the higher frequency allele tends to decrease, although to a far lesser extent. The consequence of this is that the M.P.R. tends to increase with T, even for T=5 a significant increase has been obtained. Clearly the optimal value of T will depend to some extent on the genetic parameters of the population and in particular on the population size N. However it does appear from these results that a quite low value of T may still give a reasonable increase in the mean response. Further simulation has been done for the case T=10, this being taken as the largest value which might reasonably be used in practise and yet might still give appreciable advantages over the case T=0. Comparisons of T=10 and T=0 were made for the following situations:

i)	Chromosome	A	B	C	D
	Effects	α	α	α	α
	Frequency in $Pop^n.1$	q	0	q	0
	Frequency in $Pop^n.2$	0	q	0	q
	Recombination fraction	c	c	c	

for a) Nc=0 and b) Nc=0.3125

Results are shown in Figure 6.17 with mean chance of fixation and M.P.R. plotted against q. When Nc=0 there are no differences at all as would be expected from the two locus model.



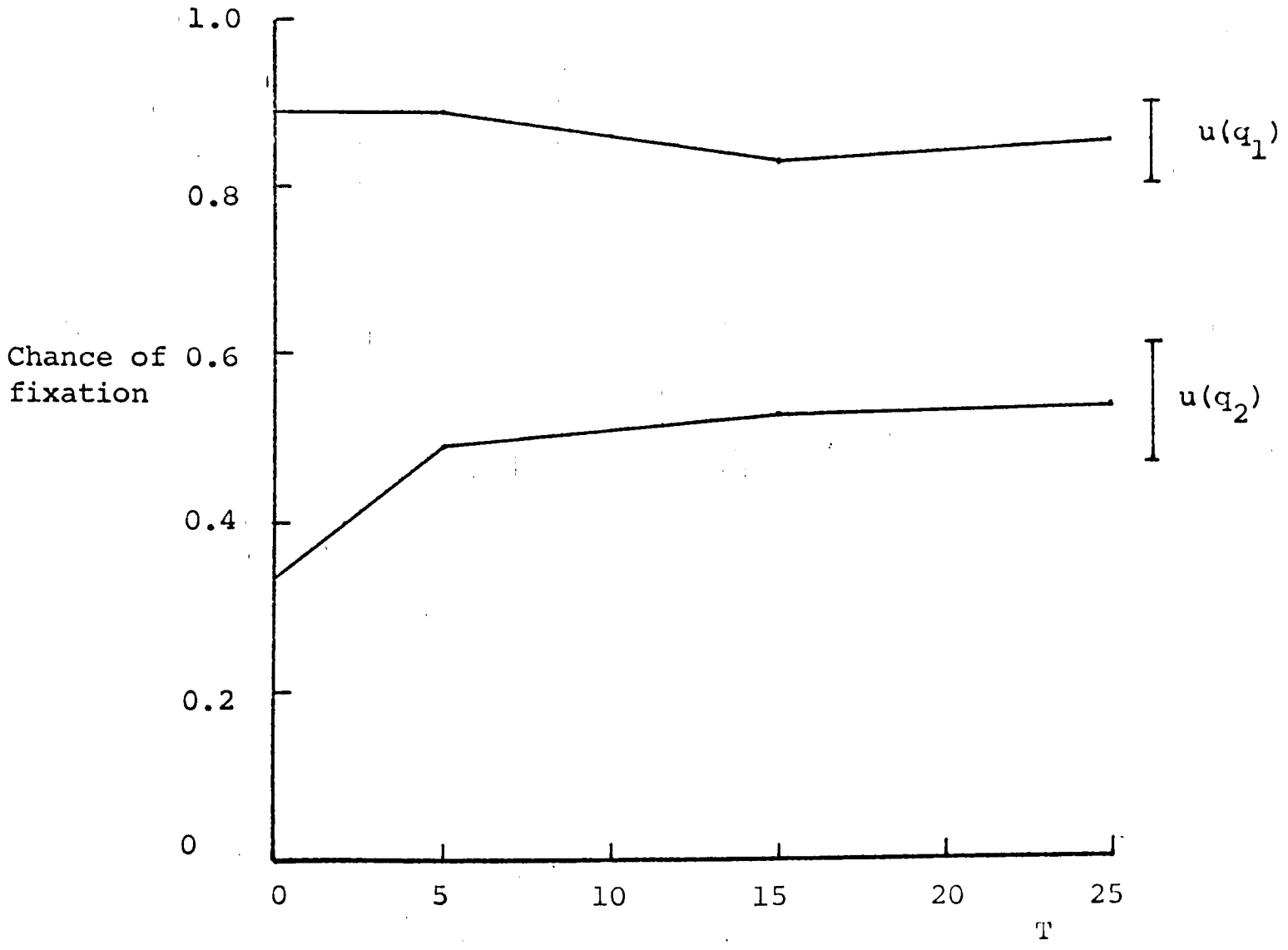


FIGURE 6.16. The effect of the length of the period of sub-line selection prior to crossing for N=10, n=4, Niα=5, Nc=0.25, h=0.5 for unequal frequencies q_1=0.8 and q_2=0.4 with selection from two base populations. Typical ranges of length four standard errors are also shown.

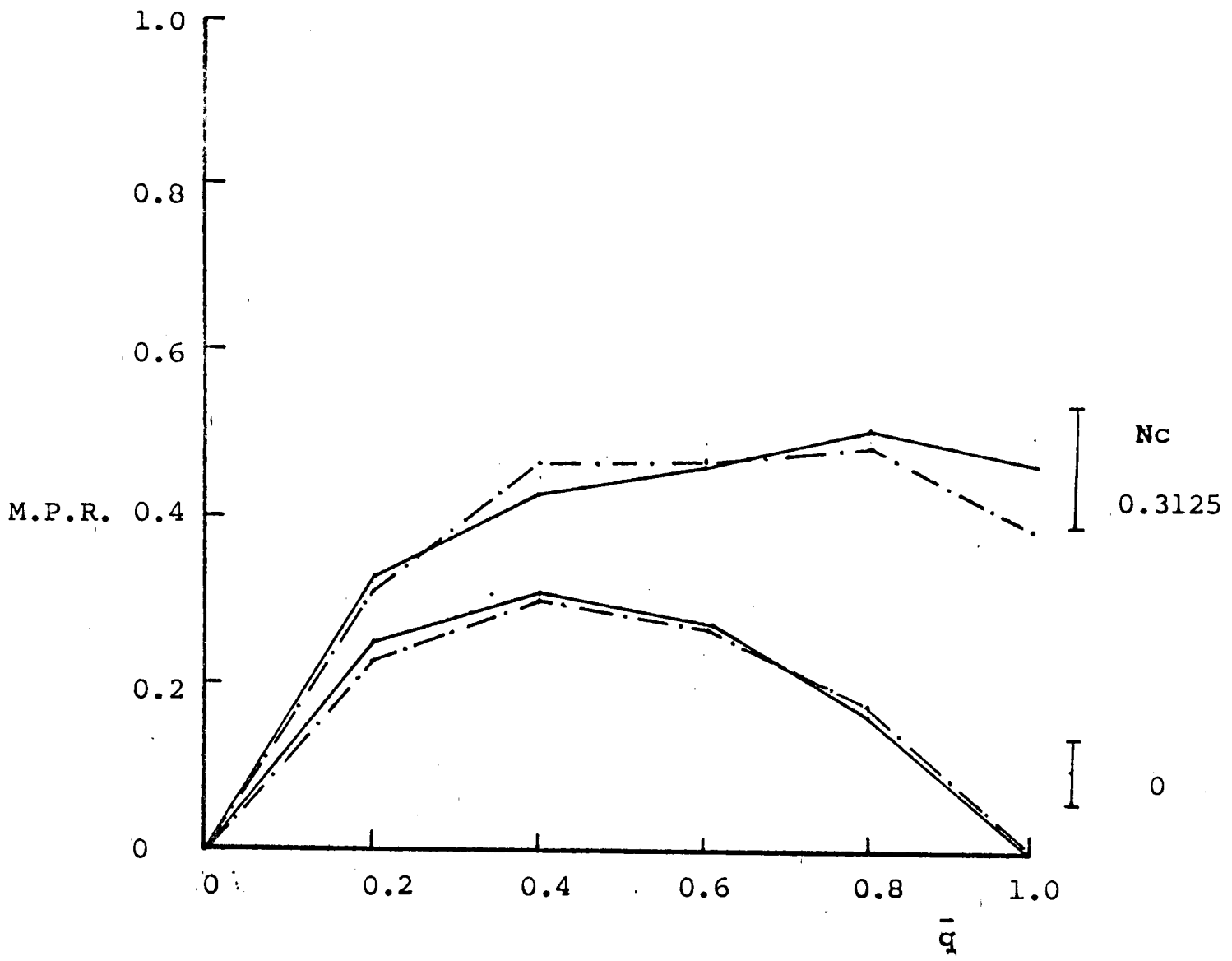


FIGURE 6.17. The effect of sub-line selection prior to crossing under a multi-locus model for N=10, n=4, Niα=5, h=0.5, T=10 for various values of $\bar{q}$ and Nc, with selection from two base populations. Typical ranges of length two standard errors are also shown.

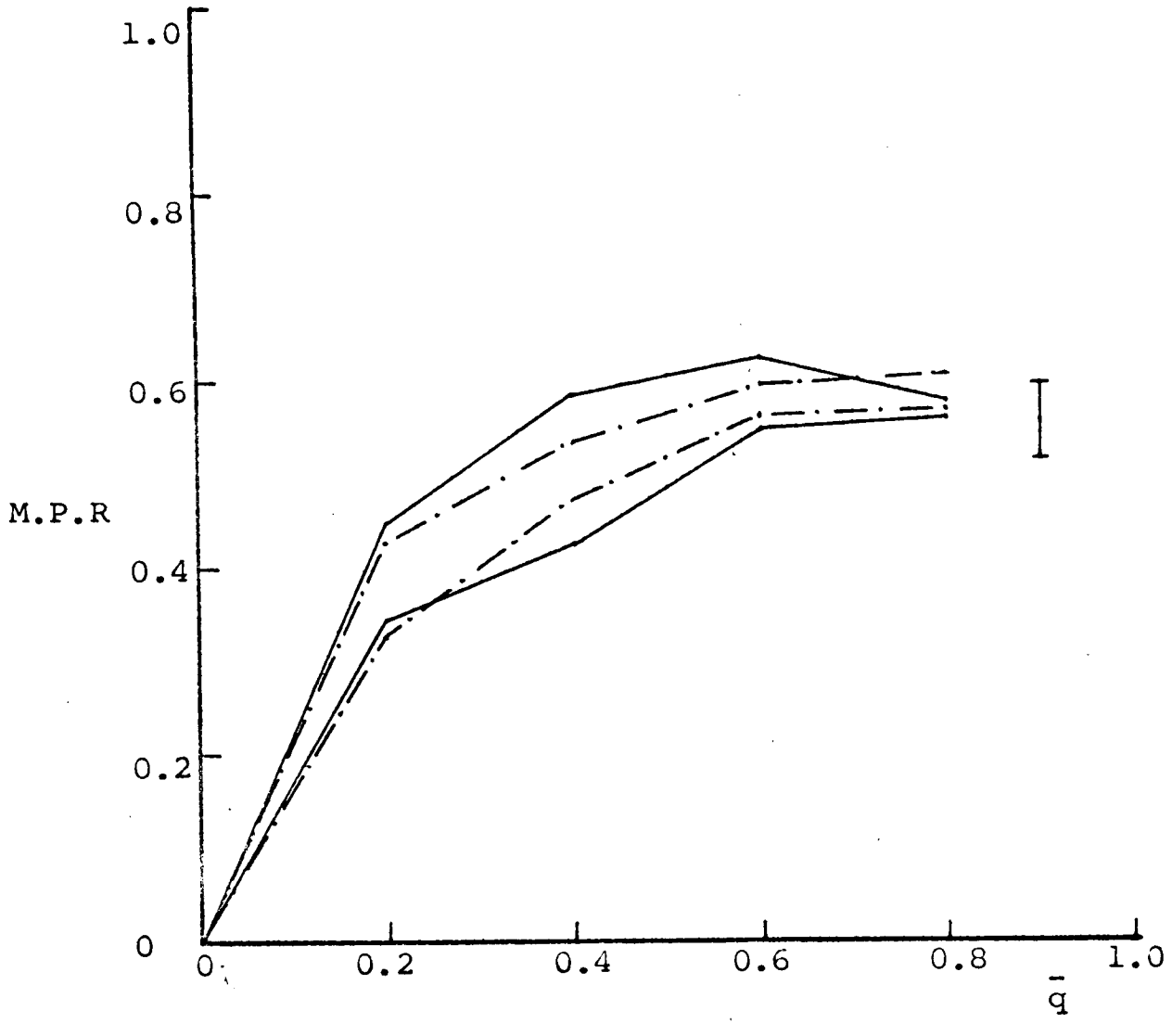
When Nc=0.3125 some differences do appear but effects of sampling are large, the only significant difference being for the case q=1.0 when clearly sub-line selection cannot make any difference. This suggests that if the mean frequencies are the same in each population differences are unlikely to be large.

ii)	Chromosome	A	B	C	D
	Effects	α_1	α_2	α_1	α_2
	Frequency in Pop^n 1	q	0	q	0
	Frequency in Pop^n 2	0	q	0	q
	Recombination fraction		c	c	c

for a) Nc=0 and b) Nc=0.3125

Results are shown in Figure 6.18 for $Ni\alpha_1 = 7.5$, $Ni\alpha_2 = 2.5$ with mean chance of fixation and M.P.R. plotted against q. For Nc=0 the larger loci gain by sub-line selection while the smaller ones lose, as a result the M.P.R. is greater under the sublining selection system although differences are not significant, this is in agreement with two locus results. For Nc=0.3125 the smaller loci still lose by the sub-line selection but in this case the larger loci also lose, this produces the reverse situation with respect to the M.P.R. which is significantly lower if sub-line selection is practised.

As has been mentioned before if one is in practice to perform selection from two separate populations in general all that will be known is the population means M1 and M2. In view of this some possible situations have been considered where the means of the two populations are constant but for different genetic



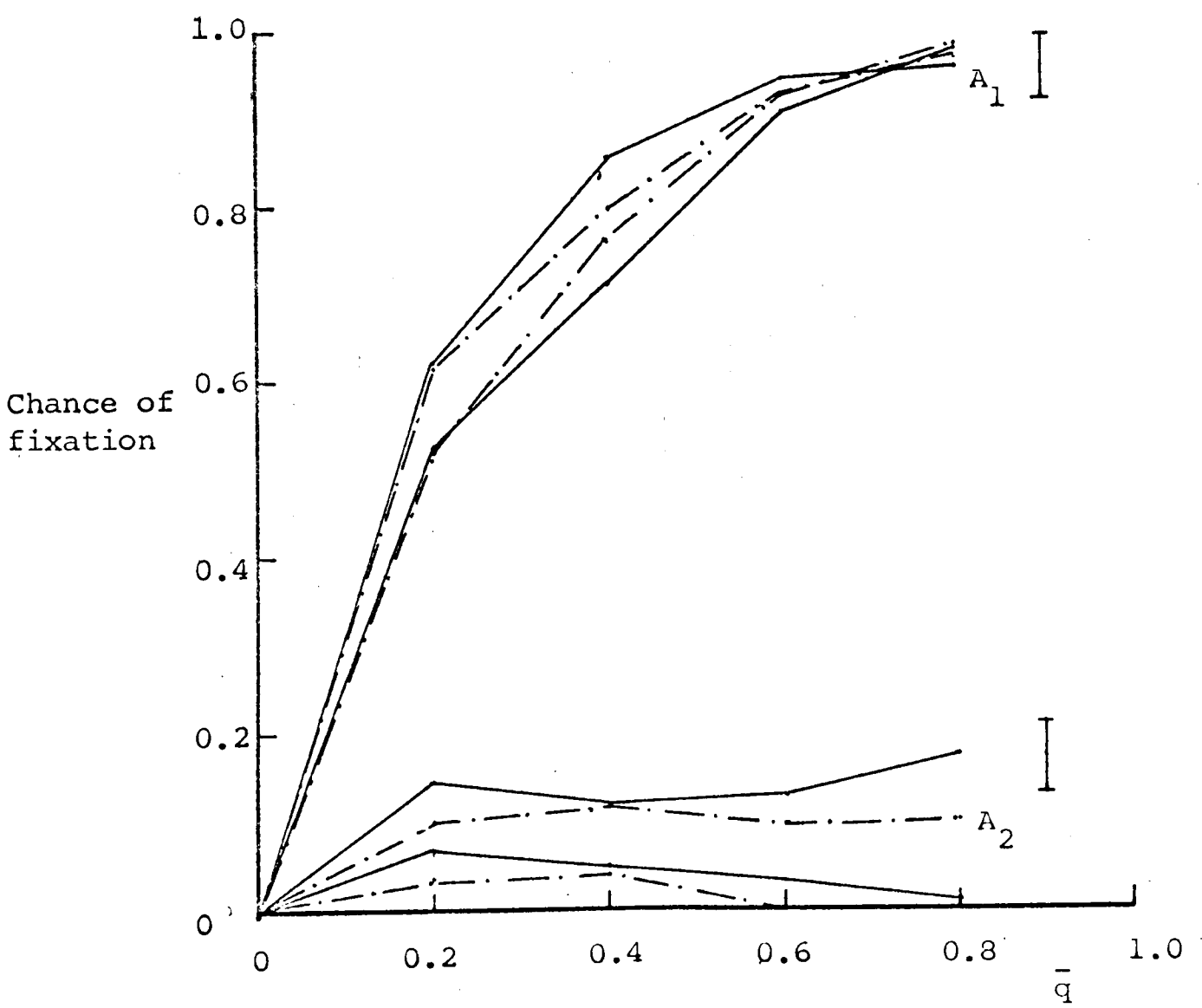


FIGURE 6.18. The effect of sub-line selection prior to crossing under a multi-locus model with unequal effects for N=10, n=4, $Ni\alpha_1$=7.5, $Ni\alpha_2$= 2.5, h=0.5, T=10 for various values of $\bar{q}$ and Nc, with selection from two base populations. Typical ranges of length four standard errors are also shown.

arrangements and the consequences of different procedures compared. This has been done for two cases:

i) M1 = M2, ii) M1 > M2

i) M1 = M2

Only situations where all four loci segregate in the cross have been considered for the situation described below:

Chromosome	A	B	C	D
Effects	α_1	α_2	α_1	α_2
Frequency in Pop^{n}1	q_{A1}	q_{B1}	q_{C1}	q_{D1}
Frequency in Pop^{n}2	q_{A2}	q_{B2}	q_{C2}	q_{D2}
Recombination fraction	c	c	c	

where $\alpha_1 = 3\alpha_2$.

The following situations have been simulated for Nc=0 and Nc=0.3125:

(a) Both populations segregating at all loci at the same frequencies

i.e. $q_{A1} = q_{A2} = q_{B1} = q_{C1} = q_{C2} = q_{D1} = q_{D2}$

(b) Both populations segregating at all loci but with differences in frequencies

e.g. $a_{A1} = q_{B1} = q_{C1} = q_{D1}$, $q_{A2} = q_{C2}$, $q_{B2} = q_{D2}$

such that $3q_{A2} + q_{B2} = 4q_{A1}$

(c) Both populations segregating at only one of each type of locus (the others being fixed unfavourably) with all frequencies otherwise equal.

e.g. $q_{A1} = q_{B1} = q_{C2} = q_{D2}$, $q_{A2} = q_{B2} = q_{C1} = q_{C2} = 0$

(d) Both populations segregating at only one of each type of locus, (the others being fixed unfavourably) with differences in frequency.

e.g. $q_{A1} = q_{B1}$, $q_{C2} \neq q_{D2}$ $q_{A2} = q_{B2} = q_{C1} = q_{D1} = 0$

such that $3q_{C2} + q_{D2} = 4q_{A1}$

(e) Both populations segregating at only two loci such that one carries large loci only and the other small loci only

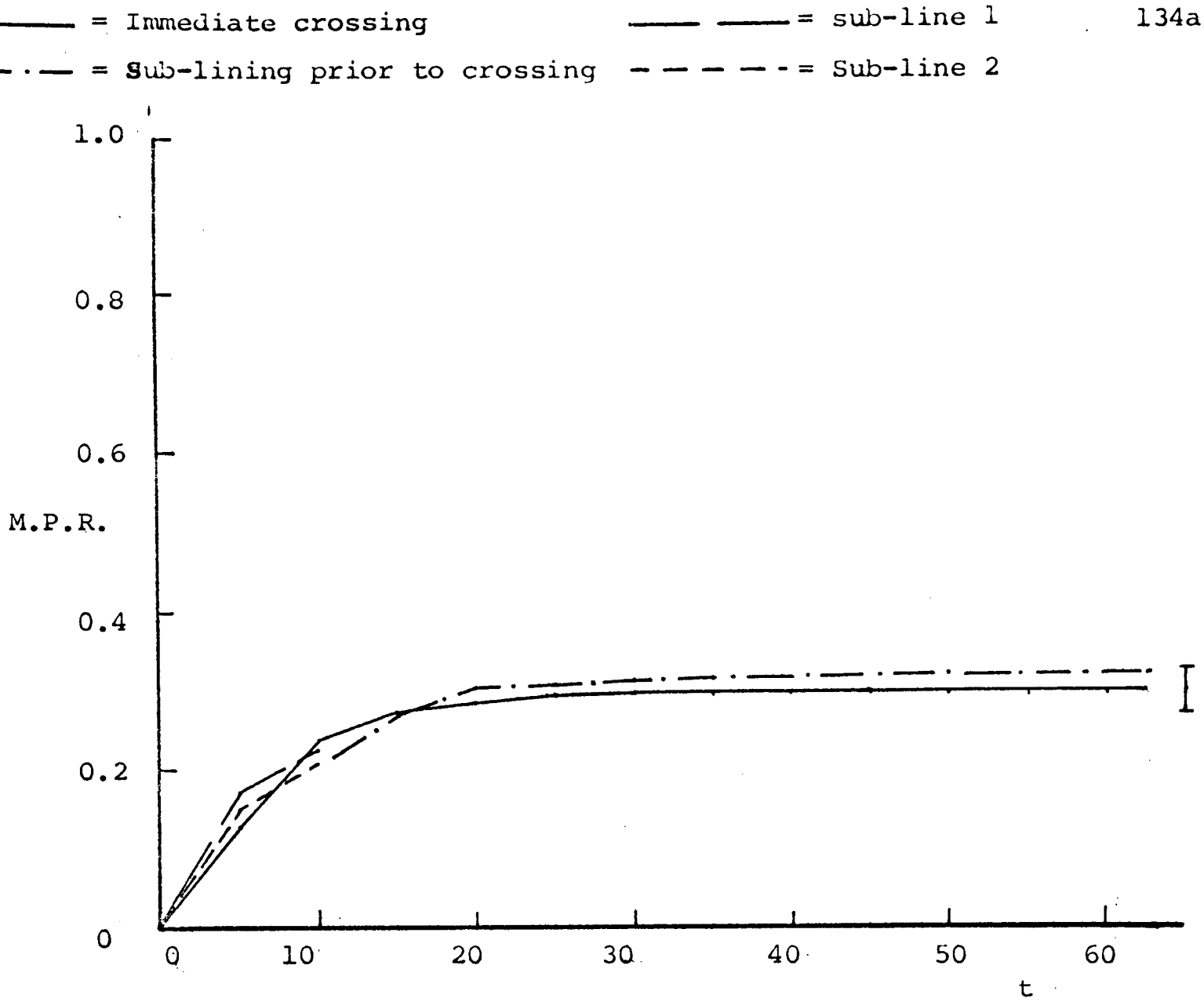
e.g. $q_{A1} = q_{C1}$, $q_{B2} = q_{D2}$, $q_{A2} = q_{B1} = q_{C2} = q_{D1} = 0$

such that $q_{B2} = 3q_{A1}$

Simulation results for M1 = M2 = K + 0.4 are set out in Table 10.

With Nc = 0 no differences significant at the 5% level are found although there is a general tendency for the M.P.R. to be greater at the limit if the sub-populations are selected prior to crossing. Results for situations (c) and (e) are shown in Figure 6.19 giving the M.P.R. plotted against t, the number of generations of selection. With Nc=0.3125, the difference for case (d) was found to be statistically significant although those for situations (c)and (e) were also quite large. In general it was again found that the M.P.R. at the limit was greater for the sub-line selection system. Results for situations (c) and (e) are again plotted as the M.P.R. for each generation of selection, see Figure 6.20.

It can also be seen from Figures 6.19 and 6.20 that the M.P.R. in the early generations is approximately the same for the two systems but as the sub-lines become fixed and the cross is made so the M.P.R. for the sub-line system falls below that for the immediate cross, and it is not until after some 20 or 25 generations that the former system shows its advantage.



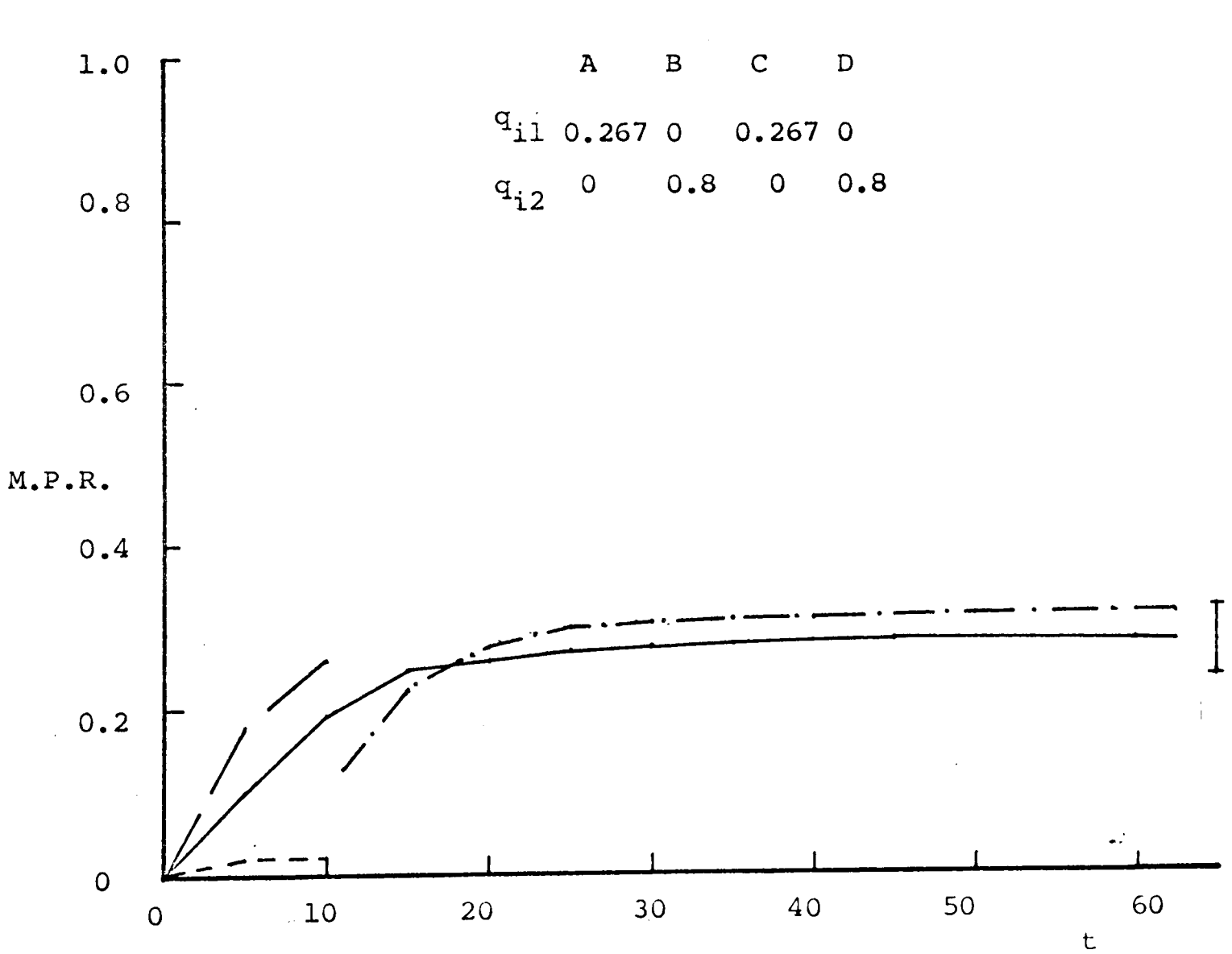


FIGURE 6.19. The rate of response under single line and sub-line selection systems for N=10, n=2, $Ni\alpha_1$=7.5, $Ni\alpha_2$=2.5, Nc=0, M1=M2=0.4, h=0.5 with selection from two base populations. Typical range of length four standard errors is also shown.

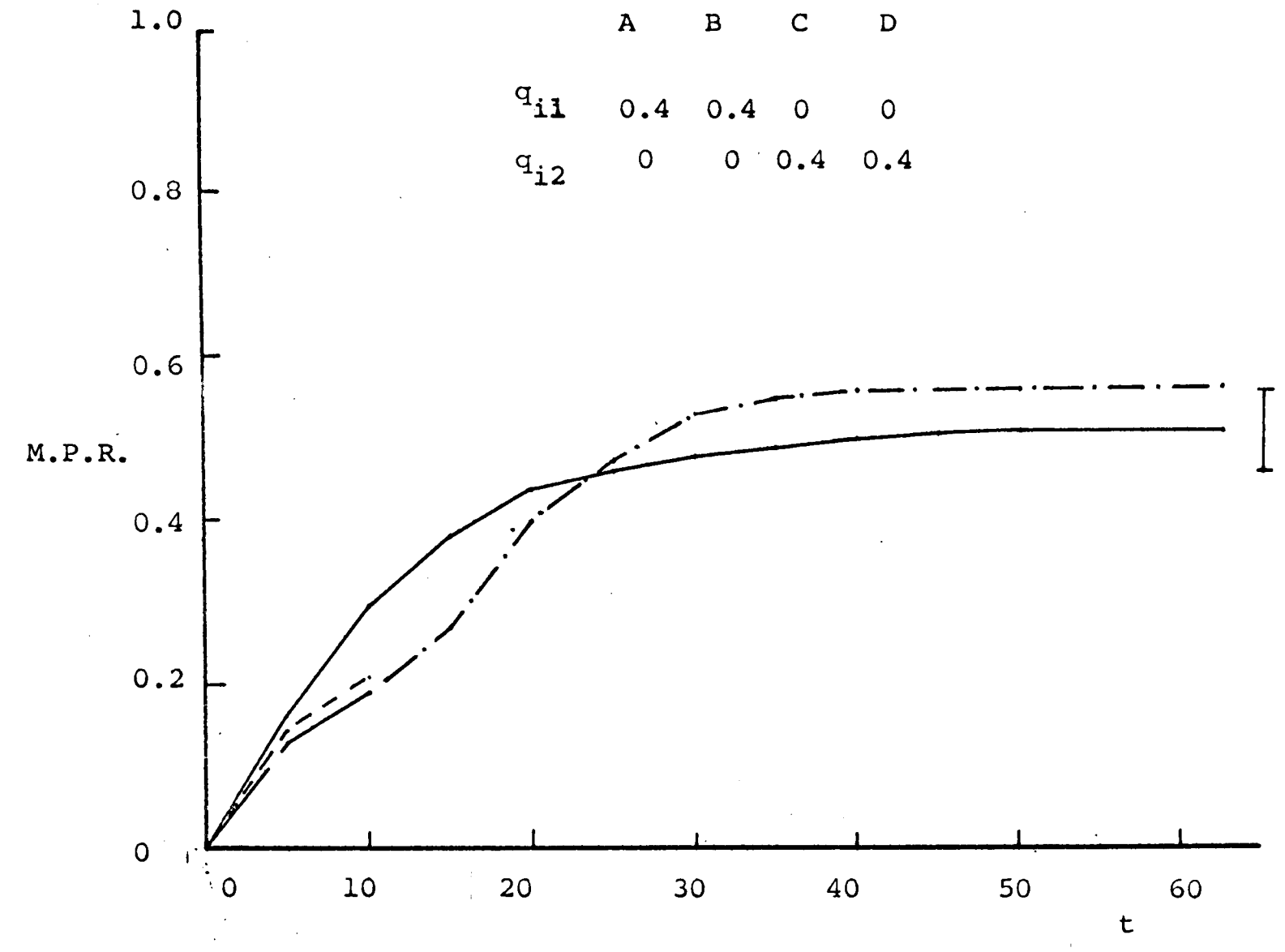


FIGURE 6.20 The rate of response under single line and sub-line selection systems for N=10, n=2, $Ni\alpha_1$=7.5, $Ni\beta_2$=2.5, Nc=0.3125 M1=M2=0.4, h=0.5 with selection from two base populations. Typical range of length four standard errors is also shown.

TABLE 10 M1 = M2 = 0.4 + K

		Chromosome				Nc	M.P.R.		
		A α_1	B α_2	C α_1	D α_2		sub-line	Single line	Difference
a	q_1	0.2	0.2	0.2	0.2	0	0.5015	0.5109	-0.0094
	q_2	0.2	0.2	0.2	0.2				
	q_1	0.2	0.2	0.2	0.2	0.3125	0.5904	0.6260	-0.0356
	q_2	0.2	0.2	0.2	0.2				
b	q_1	0.2	0.2	0.2	0.2	0	0.4109	0.4562	-0.0453
	q_2	0.1	0.5	0.1	0.5				
	q_1	0.2	0.2	0.2	0.2	0.3125	0.5852	0.5336	0.0516
	q_2	0.1	0.5	0.1	0.5				
c	q_1	0.4	0.4	0	0	0	0.3245	0.3030	0.0215*
	q_2	0	0	0.4	0.4				
	q_1	0.4	0.4	0	0	0.3125	0.5625	0.5119	0.0506
	q_2	0	0	0.4	0.4				
d	q_1	0.4	0.4	0	0	0	0.3234	0.3062	0.0172
	q_2	0	0	0.3	0.7				
	q_1	0.4	0.4	0	0	0.3125	0.5781	0.4833	0.0948**
	q_2	0	0	0.3	0.7				

cont...

Table 10 cont....

e	q_1	0.267	0	0.267	0	0	0.3175	0.2855	0.0320
	q_2	0	0.8	0	0.8				
	q_1	0.267	0	0.267	0	0.3125	0.4426	0.3833	0.0593*
	q_2	0	0.8	0	0.8				

N=10, $N i\alpha = 7.5$, $N i\alpha_2 = 2.5$, h=0.5, T=10

* Significant at 10% level

** Significant at 5% level

ii) M1 > M2

For the same chromosome arrangements the following situations have been simulated such that M1 = K + 0.8, M2 = K + 0.4.

(a) Both populations segregating at all loci with differences due to all frequencies being doubled in the first population, i.e. $q_{A1} = q_{B1} = q_{C1} = q_{D1}$, $q_{A2} = q_{B2} = q_{C2} = q_{D2}$

such that $q_{A1} = 2_{A2}$

(b) Both populations segregating at all loci with differences due to the reduced frequency of the larger loci in the second population.

e.g. $q_{A1} = q_{B1} = q_{B2} = q_{C1} = q_{D1} = q_{D2}$, $q_{A2} = q_{C2}$

such that $q_{A1} = 3q_{A2}$

(c) Both populations segregating at only one of each type of locus with frequencies being doubled in the first population.

e.g. $q_{A1} = q_{B1}$, $q_{C2} = q_{D2}$, $q_{A2} = q_{B2} = q_{C1} = q_{D1} = 0$

such that $q_{A1} = 2q_{C2}$

(d) Both populations segregating at only one of each type of locus with differences due to a reduced frequency of the larger locus in the second population.

e.g. $q_{A1} = q_{B1} = q_{D2} \neq q_{C2}$, $q_{A2} = q_{B2} = q_{C1} = q_{D1} = 0$

such that $q_{A1} = 3q_{C2}$

(e) Both populations segregating at only two loci, such that the first carries large loci only while the second carries small loci only

e.g. $q_{A1} = q_{C1}$, $q_{B2} = q_{D2}$, $q_{A2} = q_{B1} = q_{C2} = q_{D1} = 0$

such that $q_{A1} = \frac{2}{3} q_{B2}$

Simulation results are set out in Table 11 with Nc=0

TABLE 11

M1 = 0.8 + K M2 = 0.4 + K

		Chromosome				Nc	M.P.R.		
		Aa_1	Ba_2	Ca_1	Da_2		Sub-line	Single line	Difference
a	q_1	0.4	0.4	0.4	0.4	0	0.6321	0.6053	0.0268
	q_2	0.2	0.2	0.2	0.2				
	q_1	0.4	0.4	0.4	0.4	0.3125	0.7546	0.6790	0.0756**
	q_2	0.2	0.2	0.2	0.2				
b	q_1	0.4	0.4	0.4	0.4	0	0.6108	0.5983	0.0125
	q_2	0.133	0.4	0.133	0.4				
	q_1	0.4	0.4	0.4	0.4	0.3125	0.6971	0.6629	0.0342
	q_2	0.133	0.4	0.133	0.4				
c	q_1	0.0	0.8	0	0	0	0.2661	0.2582	0.0079
	q_2	0	0	0.4	0.4				
	q_1	0.8	0.8	0	0	0.3125	0.6558	0.5490	0.1068**
	q_2	0	0	0.4	0.4				
d	q_1	0.8	0.8	0	0	0	0.2710	0.2750	0.0040
	q_2	0	0	0.267	0.8				
	q_1	0.8	0.8	0	0	0.3125	0.6027	0.5567	0.0460
	q_2	0	0	0.267	0.8				

CONT....

Table 11 cont.

e	q_1	0.533	0	0.533	0	0	0.5027	0.4360	0.0661*
	q_2	0	0.8	0	0.8				
	q_1	0.533	0	0.533	0	0.3125	0.6478	0.600	0.0478
	q_2	0	0.8	0	0.8				

N=10, $Ni\alpha_1$= 7.5, $Ni\alpha_2$ =2.5, h=0.5, T=10.

* Significant at 10% level

** Significant at 5% level

no differences significant at the 5% level have been found although in all cases the sub-line selection prior to crossing gives a higher M.P.R. at the limit. Results for situations (c) and (e) are graphed as before in Figure 6.21. For the case when Nc=0.3125 some other comparisons have been made. Since M1 > M2 the possibility of a) selecting only from population 1, or b) selecting a higher proportion from population 1, i.e. putting h > 0.5 has been considered. Results are shown in Table 12, and they reveal again a result noted from the independent loci studies, that if a value of h other than 1 or 0 is to be used an intermediate value gives the highest response. In Table 12 the situations (c), (d) and (e) where different loci were segregating in the two populations a cross was always preferable to selecting only one population and then h=0.5 always gave a higher M.P.R. at the limit than h=0.7. Figure 6.22 shows the M.P.R. for situation (b) with h=1, 0.7 and 0.5 while figure 6.23 shows the M.P.R. for situation (c) with h= 1, 0.7 and 0.5. These results in general show that a higher M.P.R. at the limit may be attained if sublines are selected for some generations prior to crossing, however it is again evident that this system leads to a marked reduction of the M.P.R. in intermediate generations.

(c) Discussion and Conclusions

Returning to the questions raised at the beginning of this chapter some answers can be tentatively put forward on the basis of the work outlined above.

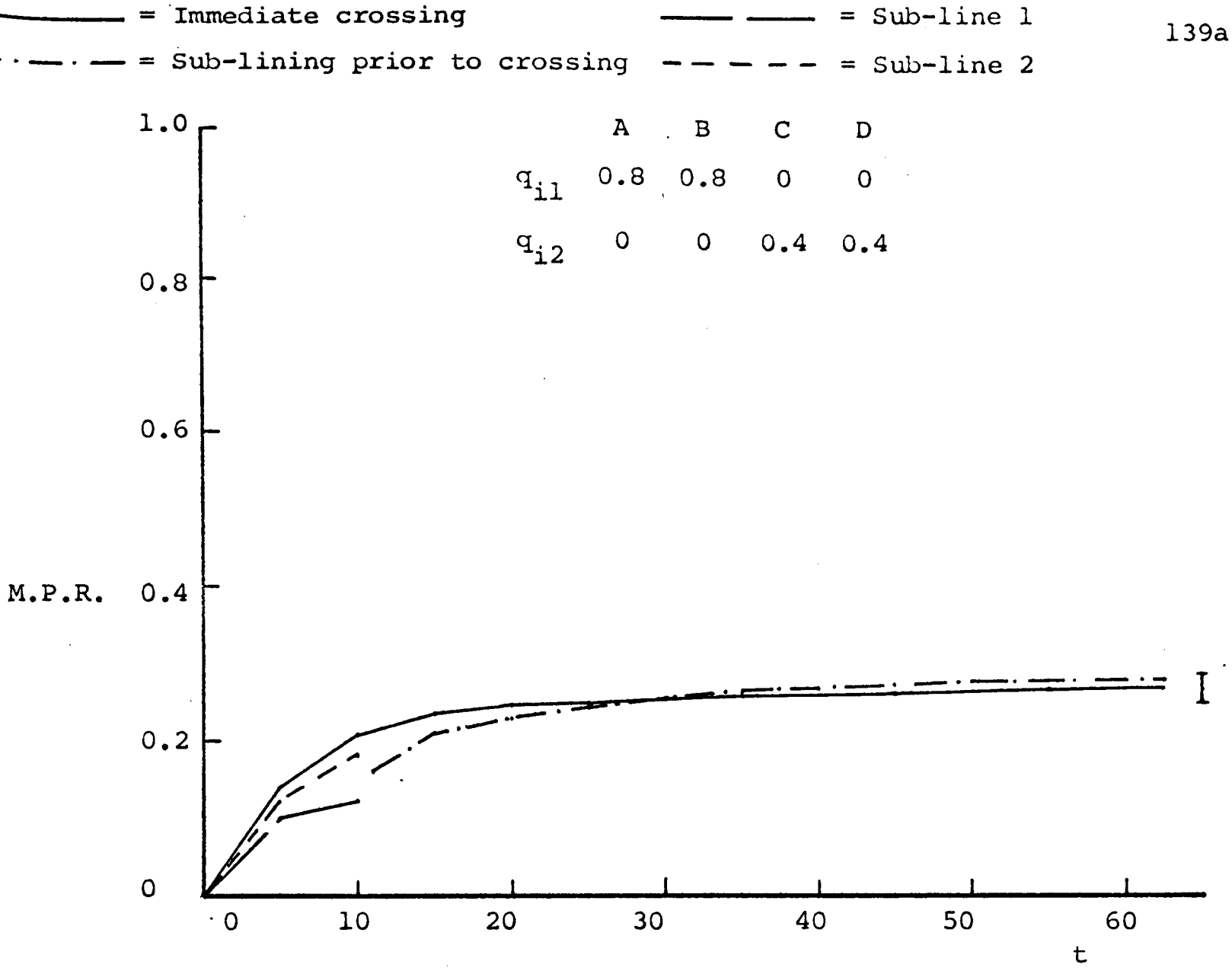


FIGURE 6.21. The rate of response under single line and sub-line selection systems for N=10, n=2, $Ni\alpha_1$=7.5, $Ni\alpha_2$=2.5, Nc=0, M1=0.8, M2=0.4, h=0.5 with selection from two base populations. Typical range of length four standard errors is shown.

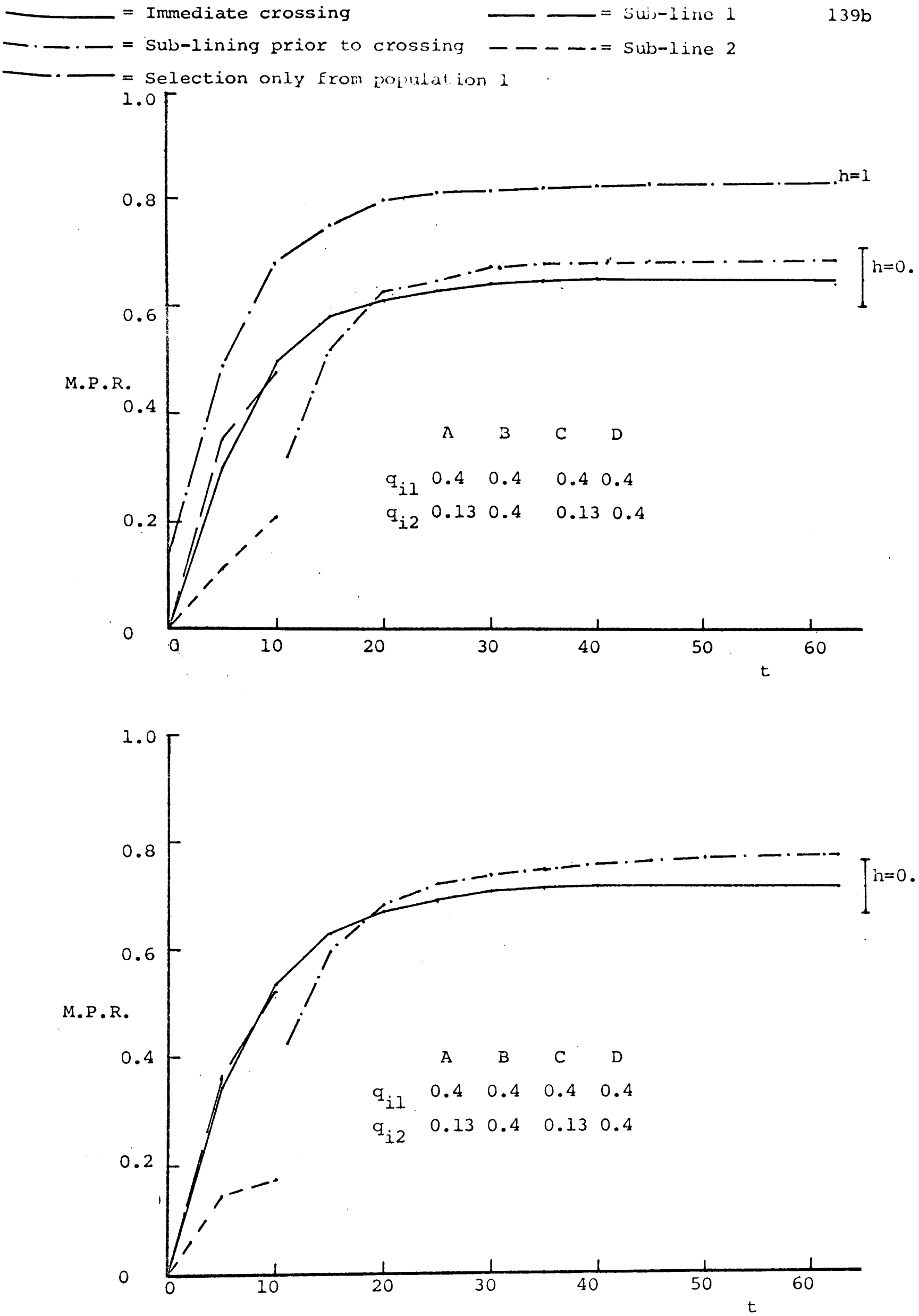


FIGURE 6.22. The rate of response under single line and sub-line selection systems for N=10, n=2, $Ni\alpha_1$=7.5, $Ni\alpha_2$=2.5, Nc=0.3125, M1=0.8, M2=0.4 with different values of h shown; selection from two base populations. Typical range of length four standard errors is also shown.

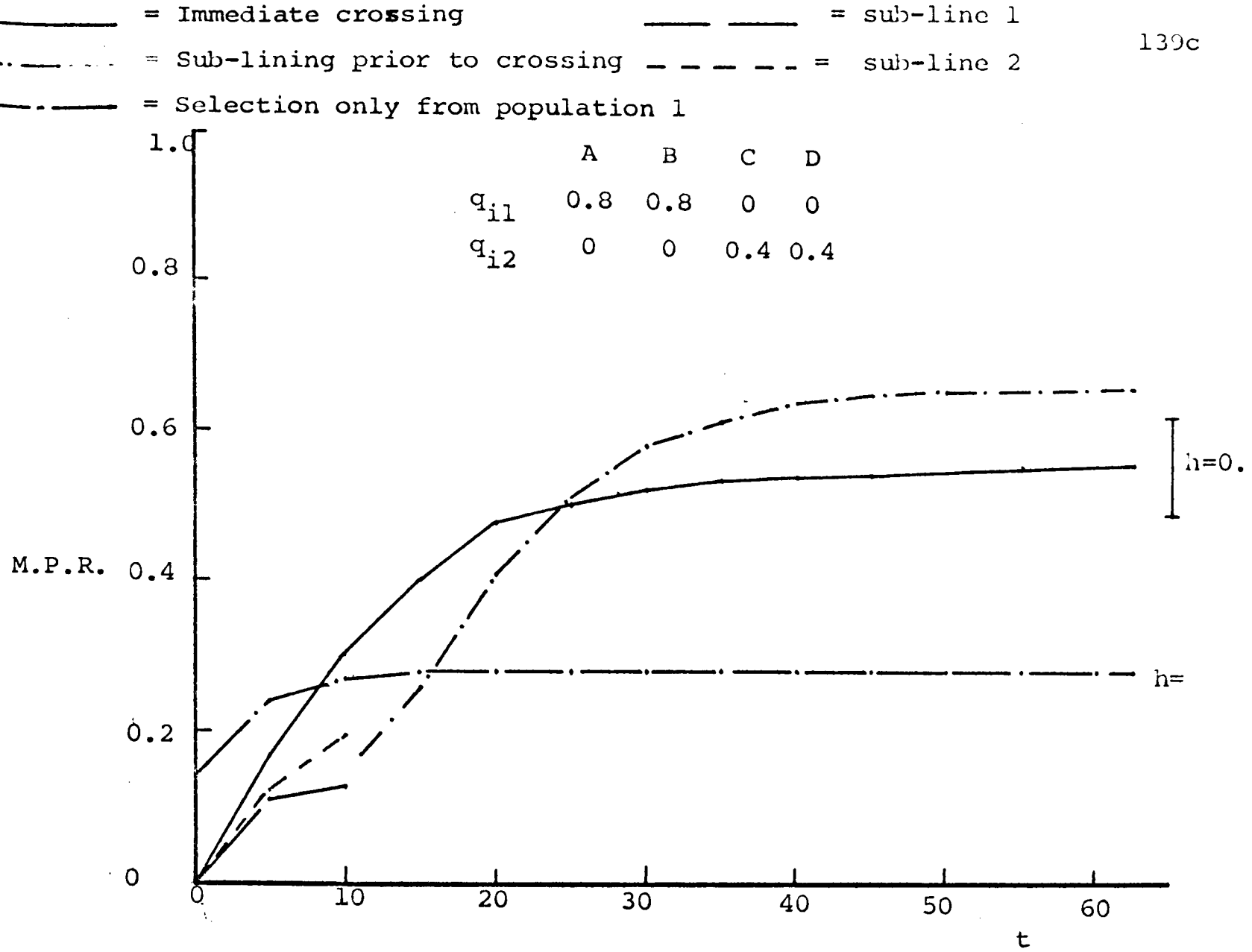


FIGURE 6.23. The rate of response under single line and sub-line selection systems for N=10, n=2, $Ni\alpha_1$=7.5, $Ni\alpha_2$=2.5, Nc= 0.3125, M1=0.8, M2=0.4 with different values of h shown; selection from two base populations. Typical range of length four standard errors is also shown.

TABLE 12

Situation		M.P.R. h=1	h=0.7	h=0.5	(h=1)-(h=0.5)
(a)	Subline	0.8409	0.7676	0.7546	0.0863**
	Single	-	0.7440	0.6790	
	Difference	-	0.0236	0.0756	
(b)	Subline	0.8409	0.7801	0.6971	0.1438**
	Single	-	0.7216	0.6629	
	Difference	-	0.0585	0.0342	
(c)	Subline	0.2812	0.4821	0.6558	-0.3746**
	Single	-	0.4712	0.5490	
	Difference	-	0.0103	0.1068	
(d)	Subline	0.2812	0.4731	0.6027	-0.3215**
	Single	-	0.4683	0.5567	
	Difference	-	0.0048	0.0460	
(e)	Subline	0.6352	0.6114	0.6478	-0.0127*
	Single	-	0.6093	0.6000	
	Difference	-	0.0021	0.0478	

$N = 10$, $Ni\alpha_1 = 7.5$, $Ni\alpha_2 = 2.5$, $Nc = 0.3125$, $T=10$

* Significant at 10% level

** Significant at 5% level

i) Should only one population be used, if so on what basis should such a choice be made?

Any such decision must be made on the basis of the population means and to some extent their variances. If the means are equal then for all cases considered under the single locus theory the cross is superior to randomly choosing one or other population. If the population means are not equal then in some cases the choice of the population with the higher mean will give the greater response, e.g. if the two populations have the same genes present but one has them at a higher average frequency. On the other hand there are some cases when a cross will give a higher response, for example if each population carries genes which the other does not have. In this case significant differences in variance may be found giving some indication of the genetic make-up of the populations but in general this will not be known. My colleague Mr Lopez-Fanjul has been working on selection problems with distinct populations of Drosophila melanogaster comparing the response in the cross with that in the pure lines. For two populations, known as Kaduna and Pacific (being from Nigeria and the West Coast of America respectively) he has found evidence that the two populations carried essentially the same genes but at different frequencies. These were unselected populations, an alternative situation under which a cross might be considered is that when populations have been selected to near fixation, in that case further advance can only be made by making some cross although there is always the possibility that even after the cross has been selected it may never surpass the original fixed population.

ii) If both populations are to be used to form a base population, in what proportions should each contribute? If all genetic parameters of the population were known it might be possible to formulate an optimum value for h, however in general it has been found that for most cases a value of h in the region of 0.5 is near to the optimum. It has also been found that dominant loci are little influenced by the value of h while additive loci are rather more so and if selection is for recessives the choice of h may be important. Nevertheless putting h=0.5 may be the best one can do, this has been found to be true for additive loci at least, even for cases where the population means are unequal but a cross is to be made.

iii) If the cross is to be made, should the populations be selected as sub-lines prior to crossing? This question has been studied for additive loci under various models and it does appear that in some cases the response at the limit can be increased by sub-line selection. However, sub-line selection prior to crossing produces a reduction in the response in intermediate generations and the ultimate increase is unlikely to be large enough to compensate for this.

CHAPTER VII

Discussion

(a) Limitations of the study

The results presented in the earlier part of this work have mainly been obtained by the use of Monte Carlo simulation. This technique provides a powerful tool for the study of theoretical problems in population genetics, allowing quite complex models to be set up and examined. In fact it is the versatility of this system which in itself presents a major problem, since for any given model of population structure the possible combinations of parameters which could be studied is enormous. For example some of the simple two locus models studied here required the specification of at least 10 parameters which for say only 4 values of each gives 4^{10} (=1,048,576) possible parameter sets. In view of this it has been necessary to severely limit the scope of this study, this has been achieved by confining attention to certain specific situations of interest, and by making some restrictions which are discussed below:

(a) Parameter limitations

For some of the parameters which specify the situations under study only one value has been used for the majority of the simulation runs while for others only restricted sets of values have been specified. The reasons for and limitations of these restrictions are given below; for:

i) Population Size.

In general the size of the population has been kept extremely

small, values of N=8 or 10 being most commonly used, in order to reduce computing time. Instead of varying N the composite parameters Niα and Nc have been considered and varied by varying the values of iα and c. The validity of this practice is discussed by Hill and Robertson (1966). In situations where the actual value of N might have been of more importance other values have been used but the results confirmed that specification only in terms of Niα and Nc was still adequate.

ii) Selection intensity.

At all times the selection intensity i has been held constant at the value corresponding to 40% selected, this is true for both individual sub-lines and the total population. This limitation has been imposed simply to keep the volume of work involved within reasonable bounds. Systems of sub-line selection involving different selection intensities within lines while still keeping the total proportion selected constant presents interesting possibilities. Some work along these lines has been carried out by Hill and Madalena (personal communication) and although their results were not entirely encouraging it does seem that this might be a line of investigation worth pursuing further.

iii) Recombination fraction.

In general the extreme case of no recombination has been studied initially and in most detail and conclusions reached have been examined under a situation of intermediate recombination.

Single locus theory gives the case for entirely independent loci and so comparison of results for Nc=0 and Nc=∞ can

be made to give some insight into the effect of intermediate recombination.

For the multi-locus models the value of c between each adjacent pair of loci has been kept constant, thus representing a single chromosome with equally spaced loci. A situation which may be of more interest in practical terms is the multi-locus case with several chromosomes such that the loci form groups with tight linkage within them but being virtually independent of each other. In this situation one would expect to find no effect of linkage disequilibrium for pairs of loci on different chromosomes, although it may have considerable importance within linkage groups.

iv) Gene effects.

For much of the two locus work $Ni\alpha$ has been held constant at a value for which linkage has in general a large effect unless gene frequencies are very high. More attention has been paid to the relative magnitude of gene effects rather than their absolute value. This is also true for the multi-locus case where it has been necessary to further restrict studies to cases where effects are equal or take one of two possible values.

(b) Model restrictions

In addition to the range of values which particular parameters may take there is also a wide range of forms the models under study may take. These include both the variety of genetic models as well as the type of situation being simulated. The following in particular are of importance:

i) Genetic models.

Apart from some single locus theory this study has been confined to consideration of additive loci, In fact little appears to be known as to the effect of linkage on response to selection for non-additive loci in finite populations.

The result from single locus theory that if selection is for an independent recessive locus there is a real advantage in sub-line selection followed by crossing, may be of considerably more practical interest if it is found to apply under linkage situations. Clearly this is an area where further investigation might usefully be carried out.

ii) Number of loci.

The majority of the studies for linked loci made here have been under a two locus model. This is the smallest possible number of loci where linkage can be studied and it has been possible to make a very detailed study of the effect of linkage disequilibrium and its importance in the results observed for the sub-lining and crossing strategies. However it is obvious that in most real situations more loci may be involved and some insight into the effect of increasing the number of loci has been possible by studying a four locus model. Although this situation is still far from truly multi-locus it does serve to illustrate the way in which the importance of initial linkage disequilibrium is affected.

iii) Initial disequilibrium.

Linkage disequilibrium was studied for situations expected to arise in making crosses of either distinct populations or selected lines. Most emphasis has been given to the extreme case

in which the two populations contributing to the cross have carried favourable alleles at only half the loci such that maximum negative linkage disequilibrium was generated. Some less extreme situations have also been considered and these suggest that the effect of the disequilibrium rapidly disappears as the two populations contributing to the cross become more alike in genetic composition.

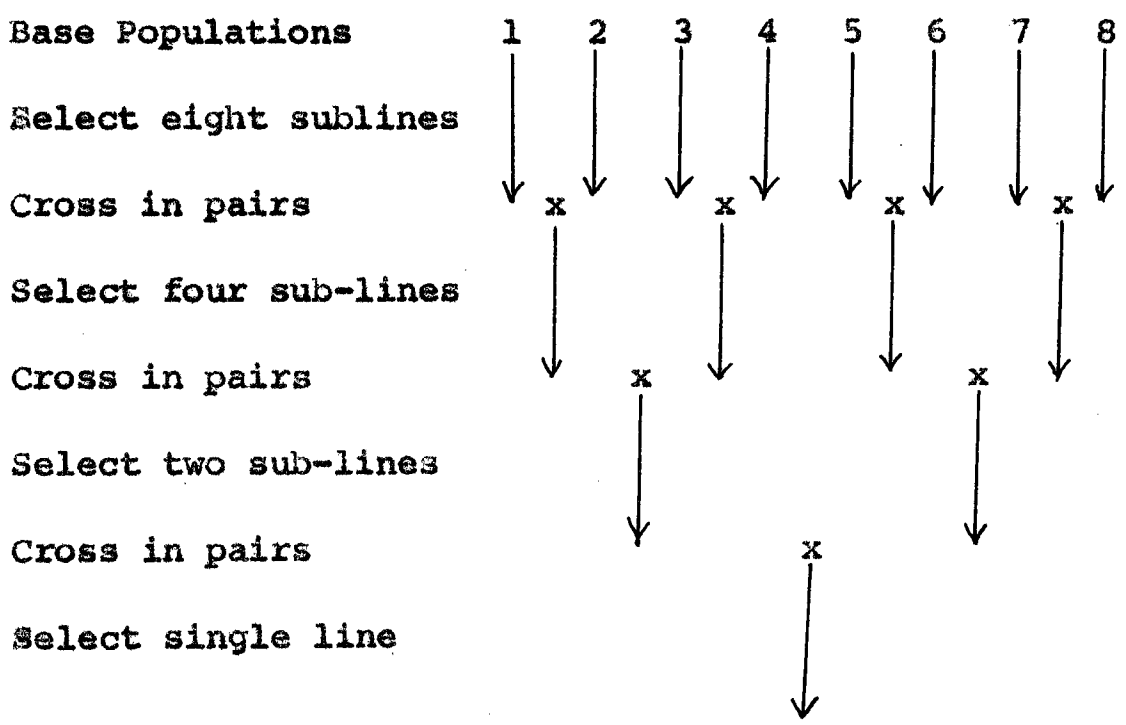
iv) Proportions in the cross.

It has been possible to make algebraic studies for the independent loci case, in which values for the proportions in the sub-lines and the cross, h and y respectively, could be formulated to give the maximum response at the limit. However, in order to find these values all parameters of the populations involved must be known, a situation which is not true in practice. In general it has been found that if two populations are to be crossed at all it is best to use an intermediate value of h or y. Therefore in most of the simulation studies made h and y have been kept constant at 0.5

v) Number of sub-lines

The number of sub-lines involved in any sub-lining and crossing scheme has been restricted to two for the present study. Theory for independent additive loci shows that chance of fixation is not changed if many, as opposed to two, sub-lines are involved. Robertson (1960) demonstrated this for the single base population case and it has been extended by Maruyama (1970) to cover the separate base populations case. The results obtained here for linked loci suggest that sub-division

of a single base population may not be affected by linkage. However if there are several base populations it may be possible to select sub-lines before crossing in such a way as to achieve a reasonable increase in the ultimate limit. For example suppose there are eight such base populations available then a selection scheme as shown diagramatically below might be of advantage



The feasibility and usefulness of such a sub-lining and crossing scheme will depend largely on the time in generations involved, clearly this requires further investigation. However in many practical instances there may be only two populations which are of a high enough standard to be considered as the basis for a selection scheme and so it is felt that the two base populations case may be of major interest.

vi) Period of sub-lining

For the two locus studies all sub-line selection has been carried out to fixation but since this is unlikely to be acceptable

in any practical programme the period has been restricted to a fixed number of generations for the multi-locus studies. For most cases the value of this has been set at 10, this being the longest time which it is thought might reasonably be used in any practical terms and yet might still allow the sub-lining to have an effect. However since rate of response is a function of population size the effect of 10 generations of sub-line selection may be quite different for a different value of N. This aspect requires further investigation if some form of sub-line selection is to be used in order to maximize the ultimate limit, particularly if economic factors are also to be taken into account.

vii) General reality of the models.

In terms of most animal population the models used in this study have been extremely simple. In all cases only two alleles at each locus have been considered and epistasis and genotype environment interactions have been entirely ignored. Sex has also been disregarded, any individual being able to mate with any other including itself. No account has been taken of natural selection and the phenotypic standard deviation, σ, has been assumed constant at all times, such that α does not change during the selection process.

(c) Computing short cuts.

As mentioned in Chapter III various short cut techniques have been used in the simulation procedure itself in order to further reduce computing time.

(b) General Conclusions

In spite of the rather formidable list of restrictions and assumptions it has been possible to draw some conclusions from this study. Since most the the results are comparative in nature, equilibrium versus disequilibrium and sub-lining versus single line selection, errors affecting absolute value have assumed less importance. The main aim has been to detect large differences, and by studying in detail situations which produce them to gain some insight into the reasons for them.

The first part of the work has been concerned mainly with the effect of initial negative linkage disequilibrium on chance of fixation and has shown that this may have a considerable effect in reducing the mean response particularly for loci with favourable alleles at low frequency or of small selective advantage. However as the number of loci involved increases from two so the importance of linkage disequilibrium is decreased. Although this is in general true the detailed studies both for two and four locus models have shown that linkage disequilibrium cannot be considered in isolation, its importance is clearly dependent on the recombination frequencies associated with it and it has also been found that the relative magnitude of the selective effects of the loci concerned are of major importance. To consider linkage disequilibrium at all under a multi-locus model (that is for more than two loci) it has been necessary to develop some appropriate form of measure which can be applied. This has been done by two rather different systems of weighting

the linkage disequilibrium, as calculated between pairs of loci, to arrive at two measures referred to here as D_{w1} and D_{w2}.

These studies on the importance of initial linkage disequilibrium were undertaken mainly in the hope of gaining some insight into the effect of linkage on some results previously derived for independent loci. In particular the result that sub-line selection followed by crossing had no effect on the ultimate selection limit of the total population size was held constant. For the case of only a single base population it was found that linkage had no appreciable effect on this result. It appears that the effect of any linkage disequilibrium generated in the cross was compensated for by the effect of linkage on the ratio of response in lines of size N and N/2. Even for situations where single locus theory was extended to show an increase in the ultimate limit for a sub-line selection system, the linked locus case gave no detectable differences. These results were not very encouraging and suggested that sub-line selection might never present a practical proposition, for although the ultimate response was the same, the sub-line system gave quite severe disadvantages in terms of response in the intermediate generations.

Rather more interesting results were obtained for the case of selection from separate base population but before presentation of these some practical considerations will be discussed. In animal breeding crossing of breeds is widely practised as a means of utilizing heterosis, pure breeds form the selected populations and a cross-bred progeny is marketed, so that

there is no cross-bred line maintained. This aspect of cross-breeding is discussed by Robertson (1971) and will not be considered in detail here. Crossing of breeds to form a new 'synthetic' breed is not a particularly new idea and it has been used with some success to incorporate different characters of merit from two breeds into a single 'new' breed. The use of synthetics to improve a single trait is less widespread although it has been practised with some success in poultry (King 1971). A problem in producing such a synthetic is finding breeds of comparable merit, most probably strains within a breed will prove to be most appropriate to use. Hill (1971a) suggests that synthetics might be produced as separate sire and dam lines such that the heterosis obtained from having cross-bred progeny can be retained. He goes on to discuss the time it might take before such a cross could be expected to surpass the better parent. This study has been mainly concerned with the effect of linkage disequilibrium in such a cross and how ultimate selection limits are affected by firstly the proportions going into the cross and secondly by sub-line selection in the separate populations prior to crossing. It has been found that in general a cross taking equal proportions from each of two sub-lines gives the highest limit in the absence of any reliable information as to the genetic composition of the populations. It has also been shown that unless considerable differences exist between populations and important loci are very closely linked then linkage disequilibrium will have very little effect on the response and sub-line selection will not

increase the ultimate limit. However if two populations contributing to a cross do differ considerably in frequency of important closely linked loci then linkage disequilibrium may have an appreciable effect and a higher limit may be attained by selecting the populations as separate sub-lines before crossing. This conclusion has been reached purely on the basis of the response at the limit and an important factor which has been ignored entirely is one of economics. Clearly if theoretical studies on selection procedures are to be of any practical value the economic aspect must be considered. It is only comparatively recently that this has been done for animal breeding programmes and a system now being widely applied is the "discounted cash flow procedure" described by Hill (1971b). This enables systems of selection to be compared at any point in time in meaningful economic terms. It may be that in such terms the gain in ultimate response found here due to sub-line selection prior to crossing, will be exceeded by losses incurred in the intermediate generations. Clearly this is an aspect of the work which could be considered further.

Another question considered in the present study has been the effect of a period of relaxation after crossing and before re-selecting. It has been found that unless linkage disequilibrium is great and linkage is reasonably tight then any realistic period of relaxation will not produce any useful increase in ultimate response. Consideration of economic factors in this instance will tend to support this conclusion and may mean that under no circumstances will waiting be of any benefit.

Perhaps the most striking fact emerging from all the results presented in this thesis is the very absence of any differences of any appreciable magnitude in spite of a variety of selection procedures involving sub-lining and crossing. It appears that provided that the values of the selection intensity, gene effects, recombination fractions, the initial mean gene frequencies and the total population size remain constant then the ultimate response remains the same. Some small differences have been found and the future of this approach to increasing selection response may be in extending these results, for example by using several base populations and selecting and crossing in some cyclical manner or as suggested elsewhere in this chapter. Alternatively selection procedures involving differences in selection intensities within lines may prove useful. Whatever might be examined next it is hoped that the results presented here will provide a useful basis for further investigations in particular with reference to the detailed study of the way in which linkage disequilibrium affects response to selection.

Acknowledgements

I am particularly pleased to thank Dr W.G.Hill who suggested the subject of this thesis and whose advice and guidance throughout has been invaluable.

I would also like to thank Professor D.S.Falconer for provision of facilities in the Institute of Animal Genetics. In addition I am grateful to many members, in particular Miss K. Paver and Mr C. Lopez-Fanjul, for help in many ways and for generally making the time spent at the Institute so enjoyable.

A research studentship from the Science Research Council is gratefully acknowledged.

Finally I would like to thank Mrs R.Gentleman for painstakingly typing the manuscript.

REFERENCES

BECKER, W.A. 1967. Manual of procedures in quantitative genetics Washington State University.

BENNETT, H.J. 1954. On the theory of random mating. Ann. Eugen. 18, 311-317.

BOWMAN, J.C. and FALCONER, D.S. 1960. Inbreeding depression and heterosis of litter size in mice. Genet. Res. Camb. 1, 262-274

BROWN, W.P. and BELL, A.E. 1961. Genetic analysis of a "plateaued" population of Drosophila melanogaster. Genetics, 46, 407-425.

CLAYTON, G.A. and ROBERTSON A. 1967. An experimental check on quantitative genetical theory. II. The long term effects of selection. J. Genet. 55, 152-170.

COMSTOCK R.E., ROBINSON, H.F. and HARVEY P.H. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. Agron.J. 41, 360-367.

CURNOW, R.N. and BAKER, L.H. 1969. Choice of population size and use of variation between replicate populations in plant breeding selection programmes. Crop. Science 9, 555-560.

DEMPSTER E.R. 1955. Genetic models in relation to animal breeding problems. Biometrics 11, 535-536.

DICKERSON G.E. 1955. Genetic slippage in response to selection for multiple objectives. Cold Spring Harb.Symp. Quant. Biol. 20, 213-224.

EWENS, W.J. 1963. Numerical results and diffusion approximations in a genetic process. Biometrika 50, 241-249.

FALCONER, D.S. and KING J.W.B. 1953. A study of selection limits in the mouse. J. Genet. 51 561-581.

FALCONER, D.S. 1960. Introduction to quantitative genetics. Oliver and Boyd, Edinburgh.

FELLER, W. 1950. Diffusion processes in genetics. Proc. 2nd Berkeley Symp. on Mathematical Stats. and Prob. 227-246.

FELSENSTEIN, J. 1965. The effect of linkage on directional selection. Genetics 52. 349-363.

FELSENSTEIN, J. 1968. Effects of linkage and epistasis on fixation probabilities. Proc. XIIth Intern. Congr. Genet. 2, 145

FISHER, R.A. 1930. The genetical theory of natural selection. Oxford University Press.

FRASER, A.S. 1957a. Simulation of genetic systems by automatic digital computer. I. Introduction. Aust.J. biol. Sci. 10, 484-492.

FRASER, A.S. 1957b. Simulation of genetic systems by automatic digital computer. II. Effects of linkage on rates of advance under selection. Aust.J. biol. Sci. 10, 492-499.

FRASER A.S., MILLER, D and BURNELL, D. 1965. Polygenic balance. Nature 206, 114

FRASER A.S. and BURNELL, D.1970. Computer models in genetics. McGraw-Hill Inc.

GEIREINGER,H, 1944. On the probability theory of linkage in Mendelian heredity. Ann. Math. Statist. 15, 25-57.

GRIFFING,B., 1960. Theoretical consequences of truncation selection based on the individual phenotype. Aust. J. biol. Sci. 13, 307-343.

HALDANE, J.B.S. 1931. A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. Proc. Camb. phil. Soc. biol. Sci. 27, 131-136.

HILL, W.G. 1965. Studies on artificial selection. Ph.D. thesis, University of Edinburgh.

HILL, W.G. and ROBERTSON, A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. Camb. 8, 269-294.

HILL, W.G. and ROBERTSON, A. 1968. Linkage disequilibrium in finite populations. Theor. and Appl. Genetics 38, 226-231.

HILL, W.G. 1969. On the theory of artificial selection in finite populations. Genetical Res. Camb. 3, 143-163.

HILL, W.G. 1971a. Theoretical aspects of crossbreeding. Ann. Genet. Sel. anim. 3(1), 23-34.

HILL, W.G. 1971b. Investment appraisal for National breeding programmes. Anim. Prod. 13, 37-50.

HULL, F.H. 1945. Recurrent selection for specific combining ability in corn. J. Am. Soc. Agron. 37, 134-145.

JACKSON N. and JAMES J.W. 1970. Comparison of three Australian Merino strains for wool and body traits. II. Estimates of between stud genetic parameters. Aust.J.Agric Res. 21, 837-856.

JAMES, J.W. 1966. Selection from one or several populations. Aust J. agric. Res. 17, 583-589.

JONES, L.P., FRANKHAM R. and BARKER, J.S.F. 1968. The effects of population size and selection intensity in selection for a quantitative character in Drosophila. II. Long term response to selection. Genet. Res. Camb. 12, 249-266.

KARLIN S. and McGREGOR J. 1968. Rates and probabilities of fixation for two locus random mating in finite populations without selection. Genetics 58, 141-159.

KIMURA M. 1955a. Solution of a process of random genetic drift with a continuous model. Proc. Nat. Acad. Sci. 41, 144-150.

KIMURA M. 1955b. Stochastic processes and gene frequencies. Cold Spring Harb. Symp. Quant. Biol. 20, 33-53.

KIMURA M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Statist. 28, 882-901.

KIMURA M. 1958. On the change of population fitness by natural selection. Heredity 12,145-167.

KIMURA M. 1963. A probability method for treating inbreeding systems, especially with linked genes. Biometrics 19, 1-17.

KIMURA, M. 1964. Diffusion models in population genetics. J. Appl. Prob. 1, 177-232.

KIMURA M. & OHTA, T. 1969a. The average number of generations until fixation of mutant gene in a finite population. Genetics 61, 763-771.

KIMURA M. & OHTA T. 1969b. The average number of generations until extinction of an individual mutant gene in a finite population. Genetics 63, 701-709.

KING J.W.B. 1971. Cross breeding of Pigs and Poultry. Proc. Xth Int. Congr. Anim. Prod. 117-130.

LATTER, B.D.H. 1965a. The response to artificial selection due to autosomal genes of large effect. I. changes in gene frequency at an additive locus. Aust. J. biol. Sci. 18, 585-598.

LATTER, B.D.H. 1965b. The response to artificial selection due to autosomal genes of large effect. II. The effects of linkage on limits to selection in finite populations. Aust. J. biol. Sci. 18, 1009-1023.

LATTER, B.D.H. 1966a. The response to artificial selection due to autosomal genes of large effect. III The effects of linkage on the rate of advance and approach to fixation in finite populations. Aust. J. biol. Sci. 19, 131-146.

LATTER, B.D.H. 1966b. The interaction between effective population size and linkage intensity under artificial selection. Genet. Res. Camb. 7, 313-323.

LATTER, B.D.H. 1969. Models of quantitative genetic variation and computer simulation of selection response. Proc. Intern. Conf. Computer Application to Genetics. Univ. Hawaii Press. 49.

LEWONTIN R.C. and KOJIMA K. 1960. The evolutionary dynamics of complex polymorphisms. Evolution 14, 458-472.

MADALENA F.E. 1970. Studies on the limits to artificial selection. Ph.D. Thesis. Edinburgh.

MARTIN, F.G. and COCKERHAM C.C. 1960. High speed selection studies. Biometrical Genetics. Ed. Kempthorne, 35-45.

MARUYAMA T. 1970. On the fixation probability of mutant genes in a sub-divided population. Genet. Res. Camb. 15, 221-227.

McPHEE, C.P. 1967. The role of recombination in artificial selection. Ph.D. Thesis, Edinburgh

MORAN, P.A.P. 1960. The survival of a mutant gene under selection. II. J. Aust. Math. Soc. 485-491.

NARAIN, P. 1970. A note on the diffusion approximation for the variance of the number of generations until fixation of a neutral mutant gene. Genet. Res. Camb. 15, 251.

NEI, M. 1963. Effect of selection on the components of genetic variance. Statistical genetics and Plant breeding. Ed. Hanson and Robinson. 501-515.

OHTA, T. 1968. Effect of initial linkage disequilibrium and epistasis on fixation probability in a small population with two segregating loci. Theor. Appl. Genet. 38, 243-248.

OHTA, T. and KIMURA M. 1969. Linkage disequilibrium due to random genetic drift. Genet. Res. Camb. 13, 47-56.

OSMAN H and ROBERTSON A, 1968. The introduction of genetic material from inferior into superior strains. Genet. Res. Camb. 12, 221-236.

POLLAK, E. 1966. On the survival of a gene in a sub-divided population. J. Appl. Prob. 3, 142-155.

ROBERTS, R.C. 1967. The limits to artificial selection for body weight in the mouse. III. Selection from crosses between previously selected lines. Genet. Res. Camb. 9, 73-85.

ROBERTSON, A. 1960. A theory of limits in artificial selection. Proc. R. Soc. B. 153, 234-249.

ROBERTSON, A. 1964. The effect of non-random mating within inbred lines on the rate of inbreeding. Genet. Res. Camb. 5, 164-167.

ROBERTSON, A 1967. Animal breeding. A. Rev. Genet. 1, 295-301.

ROBERTSON, A, 1969. The theory of animal breeding. Proc. XIIth Intern. Congr. Genet. 3, 371-377.

ROBERTSON, A. 1970a. A theory of limits to artificial selection with many linked loci. Mathematical topics in population genetics. Ed. Kojima. 246-288.

ROBERTSON A, 1970b. Some optimum problems in individual selection. Theor. Pop. Biol. 1,120-127.

ROBERTSON, A. 1971. Optimum utilization of genetic material, with special reference to cross-breeding in relation to other methods of genetic improvement. Introductory report. Proc. Xth Intern. Congr. Anim. Prod. 57-68.

SCHNELL, F.W. 1961. Some general formulations of linkage effects in inbreeding. Genetics 46, 947-957.

SMITH C. 1969. Optimum selection procedures in animal breeding. Anim. Prod. 11, 433-442.

SVED, J.A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Pop. Biol. 2. 125-141.

WATTERSON G.A. 1970. The effect of linkage in a finite random mating population. Theor. Pop. Biol. 1, 72-87.

WRIGHT, S. 1931. Evolution in Mendelian populations. Genetics 16, 97-159.

WRIGHT, S. 1933. Inbreeding and recombination. Proc. Nat. Acad. Sci. Wash. 19, 420-433.

YAMADA, Y. BOHREN, B.B. and CRITTENDEN L.B. 1958. Genetic analysis of a white leghorn closed flock, apparently plateaued for egg production. Poult. Sci. 37, 565-580.

YOUNG, S.S.Y. 1966. Computer simulation of directional selection in large populations. Genetics 53, 189-205.